4-30-2021

A climatology of tornado outbreak environments derived from unsupervised learning methods

Justin Alan Bowles
jbowles2345@gmail.com

A climatology of tornado outbreak environments derived from unsupervised learning methods

By

Justin Alan Bowles

Approved by:

Andrew E. Mercer (Major Professor)
Michael E. Brown
Christopher M. Fuhrmann
Andrew E. Mercer (Graduate Coordinator)
Rick Travis (Dean, College of Arts & Sciences)

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Professional Meteorology/Climatology
in the Department of Geosciences

Mississippi State, Mississippi

April 2021

Name: Justin Alan Bowles

Date of Degree: April 30, 2021

Institution: Mississippi State University

Major Field: Professional Meteorology/Climatology

Select Appropriate Title: Andrew E. Mercer

Title of Study: A climatology of tornado outbreak environments derived from unsupervised learning methods

Pages in Study 63

Candidate for Degree of Master of Science

Tornado outbreaks (TO) occur every year across the continental United States and are a result of various synoptic scale, mesoscale, and climatological patterns. This study looks to find what patterns exist among the various scales and how that relates to the climatology of the TOs. In order to find these patterns, principal component analysis (PCA) and a cluster analysis were conducted to differentiate the patterns of data. Four distinct clusters of TOs were found with varying synoptic and mesoscale patterns as well as distinct climatological patterns. An interesting result from this study includes the shifting of TO characteristics over time to a more synoptically forced pattern that has becoming stronger and shifted eastward from the Great Plains.

DEDICATION

I would first like to thank Dr. Andrew Mercer for inspiring this research and providing me with the baseline to get started as well as his never-ending support throughout this process as he has played a large role in making this research possible by helping me understand the material within, providing the expertise to conduct the research and correct me along the way, and overall support/confidence he has given me throughout this process. Next, I would like to thank my friends and family for their unwavering support throughout the whole process and encouraging me along the way, and more specifically my parents for providing me with whatever I needed to aid me in my journey. I would also like to thank Dr. Kelsey Ellis and Dr. Sally Horn from the University of Tennessee-Knoxville for getting me interested in research and building confidence in myself and the knowledge I gained from the various meteorology/climatology classes I took while being a student and research assistant under their supervision. I would also like to thank my professors at Mississippi State University for providing me with the knowledge necessary to aid in my research.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Tornadoes occur throughout the year in the United States whether they occur within isolated storm systems or within larger synoptic scale systems. When they occur simultaneously across the synoptic scale, there is a greater chance that they will be a part of a tornado outbreak (TO). There have been several attempts to define what a TO is. One comes from the American Meteorological Society's Glossary of Meteorology (2000), which is defined as "multiple tornado occurrences within a single synoptic scale system". Another definition that has been used in several other papers is that a TO is signaled by 6 or more tornado events within a single synoptic scale system (Grazulis 1993). The one common theme each definition has is the requirement of the outbreak to occur within a confined area and to have a minimum number of tornadoes to occur within the outbreak area. However, there is still not a set definition for a TO.

In previous studies such as Doswell et al. (2006) and Mercer et al. (2009), TOs and non-tornado outbreaks (NTOs), an event in which wind and hail reports are the dominant mode of severe weather activity, were compared to look at the synoptic scale patterns, but only 50 events were compared for each the TOs and NTOs. The events were chosen at random and were not driven by the data itself so every TO could be accounted for. But there have also been other ways in which TOs have been classified.

One of these classifications includes three different categories of TOs including a local outbreak (around 1000 miles), a progressive outbreak (one that moves from west to east), and a

line outbreak (one that occurs within a small corridor) (Galway 1977). This helps to break down TO types just based off movement and relative size without including the severe weather parameters associated with it. A classification from Doswell et al. (2006) describes a TO as 6 tornadoes occurring within a confined area, similar to the classification brought forth by Grazulis (1993), except the latter deals with a synoptic scale system and not just a certain confined area. Shafer and Doswell (2010) also found a way to rank and classify severe weather outbreaks. Their method involved several different variables including the number of severe storm reports (wind, hail, and tornado), tornado path length, strength, long-track, and killer tornadoes, as well as fatalities overall, among others. By using these variables, they were able to create a ranking value for an individual severe weather outbreak and compare it to other outbreaks to determine the strength and overall impact of the outbreak. Another classification done by Fuhrmann et al. (2014) concludes that TOs can be defined by their strength. This includes the calculation of Fujita miles, which is the integration of the Fujita scale rating for a tornado over the distance the tornado traveled. So, for a TO a cumulative rating could be found and compared to other TOs to determine how one TO was stronger than another. One way to help classify TOs and their environments is to start by investigating the synoptic scale features and patterns involved in TOs.

Synoptic scale features

There is not one defined set of characteristics to look for when forecasting a TO. An early study showing how synoptic scale features affected tornado environments comes from Fawbush et al. (1951). They found that an upper-level jet streak that was accompanied by a low-level moisture plume was conducive for a tornado environment. Another study conducted by Rose et al. (2004), found locations positioned on the right entrance or left exit region of a jet streak are in areas where rising motion is enhanced, so they are at a higher risk to experience a TO within the

jet streak. This assumption is based off the concept of mass continuity where divergence of air particles aloft leads to convergence at the surface. They also assumed that low-level warm air advection (WAA) occurs as the mechanism to generate upward vertical motion. The same study also concluded that the left exit region was the dominant area in which significant tornadoes (rating of F/EF1 or higher) were located. The exit region of an upper-level jet streak contains an indirect circulation where in the left exit region, cold air rises and in the right exit region, warm air sinks. Mercer et al. (2012) found that when the left exit region of a jet streak is coupled with a rapid decrease in surface pressure and an increase in vertical wind shear, there is an even higher risk of a TO occurring. Another study conducted by Kelnosky et al. (2018) showed that the polar jet stream (PJ) was found to be present along with the subtropical jet (STJ) in a majority of TOs in their study. When these jets are coupled with an upper-level jet streak, the TOs were shown to become stronger and more significant. The PJ was found to be more prevalent within TOs that impact the southeastern United States.

Another frequently observed synoptic scale feature is that of an upper-level trough that is located west of the expected outbreak (Mercer et al. 2011). This provides the proper rising motion associated with differential cyclonic vorticity advection (CVA) co-located with the warm sector of an associated extratropical cyclone in areas east of the Rocky Mountains. This CVA combined with the warm air advection and moisture advection from the Gulf of Mexico allows for the potential of a TO. As mentioned before, the STJ is another key synoptic scale factor that impacts TOs. The STJ helps bring in warm, moist air and usually feeds into the warm sector of an extratropical cyclone to create an unstable atmosphere. Another study conducted by Weaver et al. (2012) found that the low-level jet (LLJ) was associated with tornadic activity across the Great Plains and in the southeastern United States. Without a consistent definition of a TO, it is

hard to identify what synoptic scale features may or may not be associated with a TO. It is important to identify these patterns to better understand and forecast TOs.

Thermodynamic Features

In addition to synoptic scale processes, mesoscale features are also frequently cited when anticipating a TO. An important factor for severe storm development, and more specifically tornadogenesis, is convective available potential energy (CAPE). This value relates how much thermodynamic energy is available within the atmosphere. CAPE can hold different values within various storm modes. Rasmussen and Blanchard (1998) found CAPE to be significantly lower in non-supercellular severe weather events relative to supercell-dominant events. They found CAPE to be higher within supercell events than the non-supercell events. The same study found that when CAPE is varied with height, especially in the lower levels, there is a higher chance for supercells to develop. When CAPE values are higher in the lower levels of the atmosphere (within the lowest 3km), the lifted condensation level (LCL) is lower, along with the level of free convection (LFC) being lower. This allows a parcel of air to accelerate higher into the atmosphere and tap into the energy quicker and over a much larger layer in the atmosphere. In another study, CAPE values were found to increase between non-supercell thunderstorms and tornadic supercell thunderstorms (Thompson et al. 2003(b)), supporting the idea that CAPE values influence the development of tornadic supercells. A later study (Thompson et al. 2008) found that seasonal CAPE variability impacts energy that is available for TOs during the cool season and it results in a well-understood seasonality of TOs. These lower CAPE values are seen taking place in TOs from the Gulf Coast to the Mid-Atlantic regions and tend to rely more on the synoptic scale environment than those during the late spring and early summer. When CAPE values are recorded, they depict areas where moisture and conditional instability are

superimposed, not where they may or may not be in the future (Doswell and Schultz 2006). So further understanding the synoptic scale patterns associated with TOs also helps forecast whether values of CAPE might be the highest in the future based on low-level wind patterns and where the CAPE resides at the time of documentation. While CAPE can be used to diagnose the instability in the atmosphere and aid in the prediction of TOs, it is not meant to be used as a standalone product. A study conducted by Shafer et al. (2009), found CAPE to be a poor discriminator between TOs and NTOs. One reason is because convection acts to remove instability in the atmosphere. So, if this convection occurs before the valid time of the TO, the amount of CAPE was reduced substantially. Another reason is because the CAPE values depend on the time of year and are lower in the South and the Midwest when compared to the Great Plains (Shafer et al. 2009 and Anderson-Frey et al. 2018). CAPE values are not usually reaching their peak until late spring and early summer and are in the Great Plains whereas they are lower in the southeastern United States during the two peaks of tornado activity in the spring and fall.

Another important parameter that affects the development of severe thunderstorms is convective inhibition (CIN). This is known as the "cap" on the atmosphere and can inhibit convection enough to where the atmosphere is not able to tap into the energy that is available. When CIN is found in the atmosphere, it acts as a layer of stability and does not allow for the rapid ascent of parcels and can inhibit the development of stronger thunderstorms (Davies 2004). One way to find the layer of CIN is to look at upper air soundings. When soundings are observed with supercell thunderstorms, there are typically lower values of CIN when a tornado is produced versus when one is not produced (Rasmussen and Blanchard 1998). However, this does not mean they cannot form. Supercell thunderstorms and tornadoes can still develop with high values of CIN present, especially during nocturnal tornadoes. CIN is one of the few severe

weather parameters that inhibits severe thunderstorm growth when it is present in the atmosphere but plays a key role in the development of TOs.

The LCL is the level at which a parcel of moist air lifted dry-adiabatically would become saturated (AMS Glossary of Meteorology 2000). It is important in determining how low a cloud will be able to form in the atmosphere. When LCL heights are observed with tornadic supercells, they are found to be lower than when compared to non-tornadic supercells where tornadic supercells have an average height around 800 meters and non-tornadic supercells average a height around 1200 meters. (Rasmussen and Blanchard 1998). According to the study, when heights begin to rise, the probability for a supercell to occur begins to decrease. LCL heights can also be detrimental to the growth of supercells if they are too high even in areas with high CAPE and high shear (Rasmussen and Blanchard 1998). If the LCL heights are low enough, it can be a good discriminator between weak and significant tornadoes when coupled with higher shear values (Grünwald and Brooks 2010). This could be the case as the lower the LCL value becomes, the greater the chance the tornado will be significant, whereas the higher the LCL value, there is a higher chance a weak tornado will occur.

Kinematic features

One of the main types of kinematic features to look for when forecasting for a TO includes storm relative helicity (SRH) and effective layer bulk shear (EBS). The National Weather Service defines helicity as a property of a moving fluid which represents the potential for helical flow to evolve. This is where a storm obtains its rotation. According to NOAA's Storm Prediction Center (SPC), SRH is a measure of the potential for cyclonic updraft rotation in supercells for both 1km and 3km above ground level. Rasmussen (2003) found 0-1km SRH to be a better forecast parameter for distinguishing the tornadic supercells and non-tornadic supercells

when compared to the 0-3km SRH. They came to this conclusion by analyzing the values of 0-1km SRH and the value of 2-3km SRH to see the differences between the two layers of SRH. They found the 0-1km SRH to be significantly larger than the 2-3km SRH on soundings of environments that produced tornadoes. Another study conducted by Thompson et al. (2003(b)) also found 0-1km SRH to be able to discriminate well between significant tornadic supercells and non-tornadic supercells.

The EBS normalizes the shear values for both shallow and taller storms, which allows for a more realistic assessment of a storm's vertical profile (Thompson et al. 2007). The effective inflow layer is at the inflow level of the storm and begins at the first level in which CAPE values are over 100 J/kg with CIN values greater than -250 J/kg. The EBS was designed to identify both surface-based and elevated supercell environments, with supercells becoming more probable as the EBS increases beyond 25-40 kt (Thompson et al. 2003(a)). According to Thompson et al. (2004), EBS tends to increase throughout the depth of a storm for both supercells and non-supercells, but the EBS is found to be stronger within the supercell thunderstorm. The study also suggests that the largest differences in the EBS between the supercell and non-supercell storms are located within the middle portion of the storm between the LFC and EL. Supercells would be affected by the higher EBS values because the horizontal vorticity rolls that are created from it and are picked up into the supercell, will be much stronger than those in areas with lower EBS values.

Composite Indices

There are a few composite indices available to forecasters that can aid in the prediction of supercells and TOs. One of those includes the supercell composite parameter (SCP). This composite includes most unstable CAPE (MUCAPE), 0-3km SRH, and the denominator of the

Bulk Richardson Number (BRN) in its calculation for this value (Thompson et al. 2003(b)). The BRN is defined by Weisman and Kemp (1986) as a nondimensional ratio of the CAPE to a measure of the vertical wind shear that is used to characterize convective-storm types for various environments. BRN values less than 45 support supercell convection while greater than 45 support multicell or ordinary cell convection. The SCP index was created to identify environments capable of supporting supercells. The index also normalizes CAPE and other measures of vertical wind shear to approximate threshold values for supercells and combine them into one single value. According to Thompson et al. (2003(b)), this value can also help differentiate between significant tornadic supercells and non-tornadic supercells. A value greater than 1 indicates there is a higher likelihood of supercells occurring whereas a value below 1 usually indicates the environment is not conducive for supercells to occur, although discrete thunderstorms are still possible.

While the SCP can help differentiate between tornadic and non-tornadic supercells, it is not necessarily built to do so, which is why the significant tornado parameter (STP) was created. It was created as a tool to aid forecasters in discriminating between significant tornadic and non-tornadic supercell environments (Thompson et al. 2002). This parameter is comprised of two different parameters that eventually give a single value for the STP. The two components include supercell and tornado components. The supercell components include the 0-6km AGL vector shear magnitude and the mixed layer CAPE (MLCAPE). The tornado components include the 0-1km SRH, mixed layer LCL (MLLCL), and mixed layer CIN (MLCIN). When these values are combined, changes in some of the values lead to an overall change in the value of the STP. The STP approaches zero as the CAPE or shear parameters approach zero. It also approaches zero as the LCL height goes to 2000 meters and above or if the CIN reaches a value of 150 J/kg.

According to Thompson et al. (2003(b)), significant tornadic supercells can be distinguished from non-tornadic supercells once the STP value crosses a threshold of 1. This index is not meant to be used as a standalone variable and its value is not always indicative of the outcome of a synoptic or mesoscale system. But it is a valuable tool to help forecast what the environment can do over a certain period.

Another valuable index that is commonly used is the energy-helicity index (EHI). This value was first developed by Hart and Korotky (1991) and is a combination of CAPE and SRH. In the calculation it includes the MLCAPE and 0-1/0-3km SRH. The EHI is used to diagnose the likelihood of a rotating convective updraft and aids in the prediction of whether the environment has the potential to produce a supercell or mesocyclone (Thompson et al. 2003(b) and Rasmussen 2003). In a study conducted by Rasmussen and Blanchard (1998), the 0-1km EHI was substantially better at distinguishing between significant tornadic supercells and non-tornadic supercells. A value greater than 1 indicates the potential for supercells to develop. Values that appear over 3 indicate the enhanced possibility of a strong mesocyclone-induced tornado to occur with a strength of F/EF2 or higher. This is becoming an increasingly useful index when forecasting supercell thunderstorms capable of producing tornadoes. EHI is like other indices in that it is to be used in association with other parameters when forecasting for severe weather.

When determining how to analyze the data within the study, it was important to find a way that lacked any predetermined biases from the investigator and non-hierarchical cluster analysis seemed to be of use for this study. Other studies such as Daoust (2013), Lee (2011), and Mercer et al. (2012) used cluster analysis to separate and characterize different synoptic scale patterns on tornado days to diagnose the synoptic conditions for those events. This led to the

identification of a few specific synoptic environments that were favorable for tornadic activity. Based upon the results from these studies it is fair to assume that this process will also work for this study. The cluster analysis provides an objective approach to grouping the data instead of a subjective approach. Cluster analysis will also allow the data to drive the outcome and will not be affected by any preconceived notions or biases the investigator might have. It also will allow us to identify underlying processes that could be unknown to the investigators existing knowledge.

This study offers an updated synoptic climatology of North American TOs. Before the physical characteristics of a tornadic environment are investigated, a proper definition of a TO is needed. However, given the diversity in possible definitions, the TO definition in this study is being replaced with an objectively established grouping of outbreaks (based on previous report data such as numbers of tornadoes and ranking index) whose impacts emulate those in a typical TO. Composite synoptic and mesoscale fields from this new set of cases will then be constructed to understand the physical character of this new grouping of events. Ultimately, this approach will yield a completely data-driven depiction of the typical TO environment.

CHAPTER II

DATA

Before clustering outbreak characteristics and identifying a typical TO environmental setup, a database of outbreaks was required. The outbreak database from Shafer and Doswell (2010) includes outbreak characteristics from multiple (over 6000) severe weather outbreaks, such as the numbers of tornadoes, wind reports, hail reports, and an outbreak intensity ranking index, spanning 1950 – 2011, all derived from the Storm Prediction Center's (SPC) National Severe Weather Database. Each outbreak was comprised of reports spanning 06Z-06Z (midnight to midnight CST) such that each outbreak was treated as a separate tornado outbreak day. The valid time of the given outbreak was centered on a 6-hour block that contained the most total storm reports after the reports were discretized into 3-hour groups. For example, if the 6-hour block from 21-03Z contained the most storm reports, the valid time for this outbreak would be 00Z. Some reports may not be included within this time frame of the outbreak (i.e., for events whose impacts continued after midnight), but the valid time rarely occurred outside of the 1800-0000 UTC time window such that the associated synoptic scale and mesoscale environments were still characterized appropriately by this time definition.

Once the valid time has been determined, the atmospheric conditions can be obtained from the National Centers for Environmental Prediction- North American Regional Reanalysis (NCEP-NARR) dataset (Mesinger et al. 2006). The NARR dataset is a regional dataset spanning

North America with a 32km resolution and 29 vertical levels, plus the surface, that will be used to develop the composites for the TOs. The domain selected for each event is set by a 121x81 grid centered on the outbreak, meaning it is large enough to contain all importance synoptic scale and mesoscale features. 120 grid points were used in the west and east directions with 80 grid points to the west and 40 grid points to the east. This was done in order to observe the patterns downstream of the TO region. The 80 grid points were used by 40 grid points to the north and 40 grid points to the south of the TO region. The remaining 1x1 grid was the actual location of the TO. Figure 1 shows a 300mb wind speed and geopotential height map of the April 27, 2011 TO. The data displayed on Figure 1 comes from the NARR dataset previously mentioned and was created using the Grid Analysis and Display System (GrADS) program.

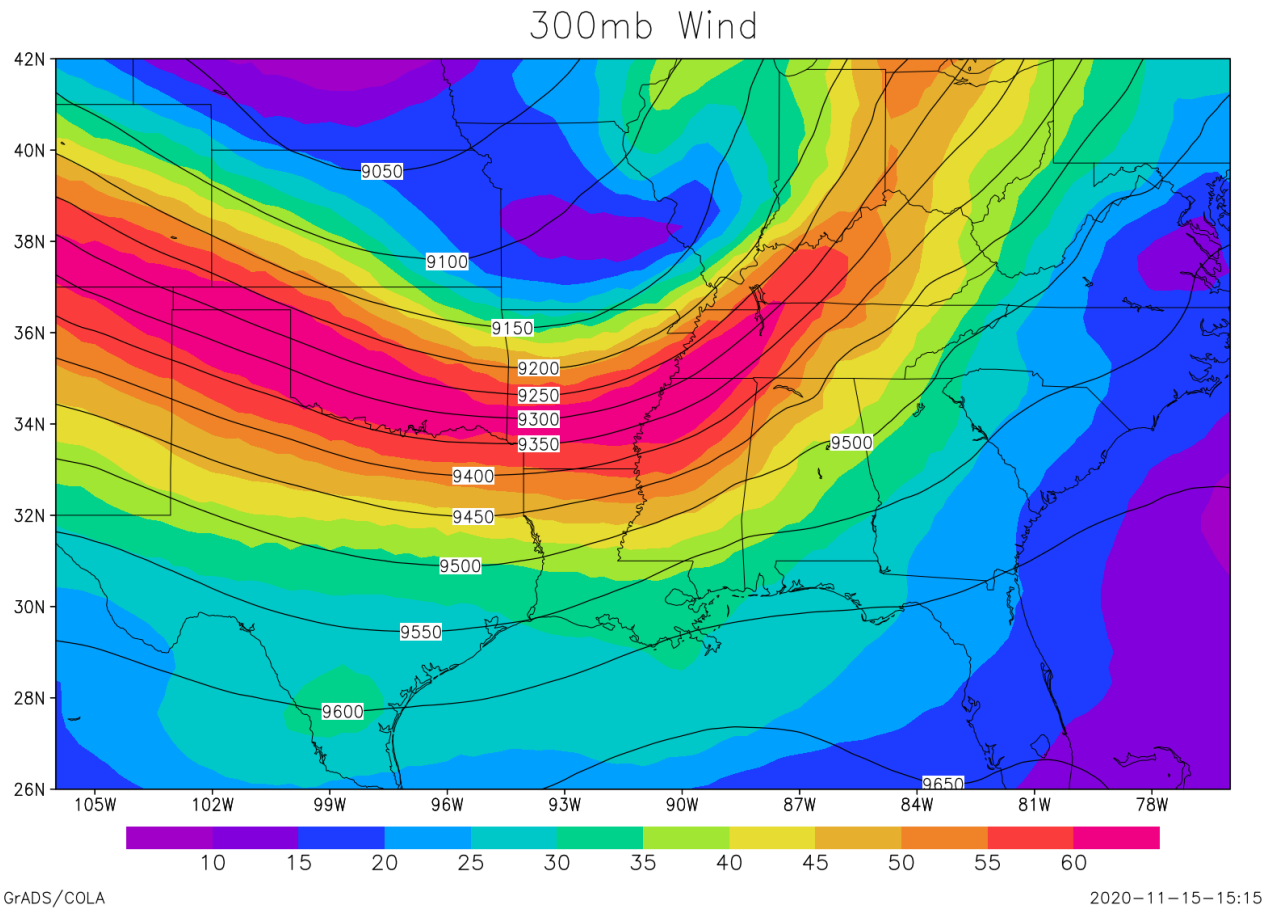

Figure 1 Sample of the extent of the domain that will be used for the composite maps. Example using 27 April 2011 Tornado Outbreak

The extent of the domain in Figure 1 is very similar to the one that will be used in the creation of the composite maps from each cluster. The data that is being used in this map will also be the same data that will be used to conduct the cluster analysis that is mentioned later in CHAPTER III.

The NARR data is available from 1979/01/01 through the present but for this study only TOs from 1979 through 2011 will be analyzed. The range of these years coincides with the

availability of the NARR reanalysis data. There are 5 meteorological variables that will be used to create the composites needed for the analysis of the data including geopotential height, specific humidity, temperature, and the *u* and *v* wind components. Once the subset of TOs were created, the NARR data was then obtained for the valid time of each TO and used within the cluster analysis.

CHAPTER III

METHODS

Cluster analysis

As stated in the introduction, a primary objective was to create a data-driven representation of a TO. To obtain this representation, a cluster analysis was performed on the storm reports for the full severe weather outbreak database, yielding different severe weather outbreak configurations. This was done by first detrending the storm reports by year to remove secular biases by transforming each severe weather outbreak's report counts to standard anomalies based on the annual mean and standard deviation of the report count for all severe weather outbreaks in that given year. The detrended Shafer and Doswell (2010) ranking index data were then Winsorized, to remove extreme outliers that would dramatically affect cluster analysis. This Winsorization procedure set any ranking index value below the 5^{th} percentile of all the values to a value of the 5^{th} percentile and repeated the procedure for the 95^{th} percentile. This showed that the long right tail in the ranking index data was a result caused by a few stronger outbreaks. After this was done, the transformed outbreak matrix was put through a simple 2 principal component (PC) principal component analysis (PCA) on the 4 variables (tornado, wind, and hail reports and the outbreak ranking index). A K-means clustering was then done on the PC scores from the PCA. Finally, since K-means analysis requires the user to estimate an appropriate number of clusters, the silhouette coefficient (Rousseuw 1987) was used

to identify the clustering that yielded the most distinct clusters. The silhouette coefficient is calculated as:

$$Silhouette(\boldsymbol{a}) = \frac{Seperation(\boldsymbol{a}) - Cohesion(\boldsymbol{a})}{\max\left[Seperation(\boldsymbol{a}),\ Cohesion(\boldsymbol{a})\right]} \quad (1)$$

In this equation, separation represents the measure of the distance between the nearest members in different clusters. The cohesion represents the average distance from the center to the member of a given cluster. It is best to maximize the cluster separation and to minimize the cluster cohesion (Mercer et al. 2012). This approach yielded 7 distinct clusters with differing severe weather outbreak characteristics in terms of storm reports and ranking index. Boxplots of these groups were then created to visualize the differences among the clusters. Of these 7 groups, cluster 4 showed to be the cluster containing the most tornado and hail reports with fewer wind reports, meaning this cluster showed the potential for storms that were of more of a supercellular nature (high hail and tornado activity but minimal straight-line wind activity) and that contained the highest-ranking index values of the 7 groups. This cluster was thus deemed the TO cluster, and outbreaks in this group were retained for analysis in this project. **Error! Reference source not found.** shows the high volume of tornado and hail reports compared to the wind reports, further confirming the likelihood of a supercellular event or TO. This cluster included 128 total outbreaks and included major outbreaks such as the 27 April 2011 Super Outbreak and the 3 May 1999 Oklahoma outbreak.

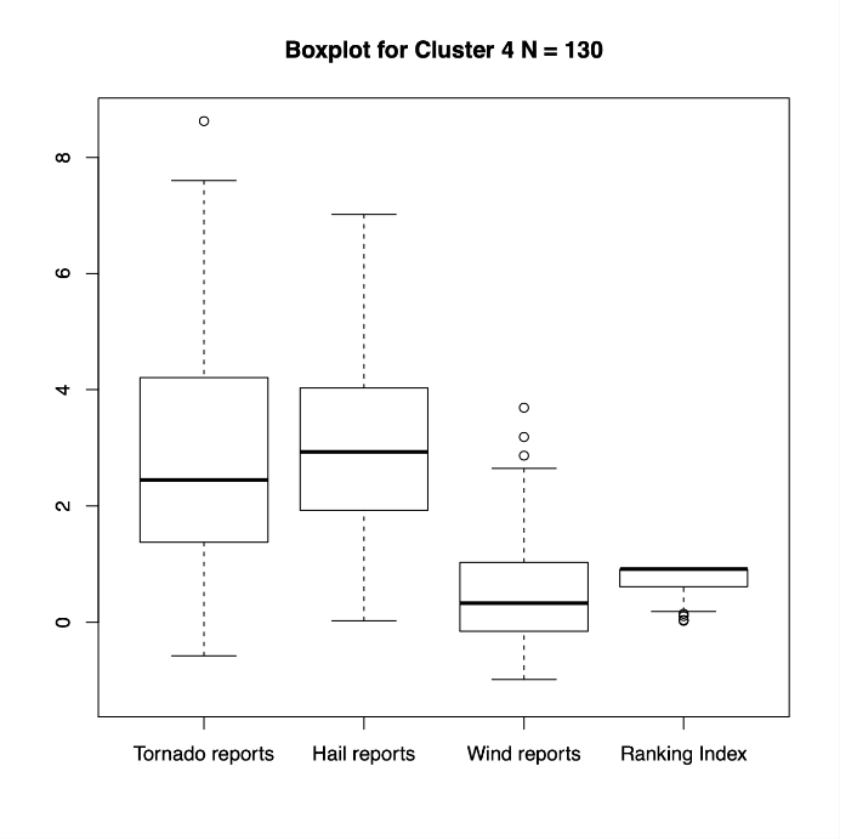

Figure 2 Cluster 4 from cluster analysis illustrating the high volume of tornado and hail reports.

Composite building

With the TO events obtained, the next phase of the work was to develop composite fields characterizing the environments of these newly defined TOs. One key statistical method that works well to capture the variability of the TOs and to develop the composites is throughout principal component analysis (PCA) (Richman 1986). The PCA equation is given as:

$$\mathbf{Z} = \mathbf{F}\boldsymbol{A}^{T} \tag{2}$$

where **Z** is the standardized anomaly matrix of the storm report dataset, **F** is the matrix of principal components (PC), and **A** is the PC loading matrix (Mercer et al. 2012). To determine **F** and **A,** the data within **Z** are standard anomalies of the original composite fields. This standardization will be done by grid point (so that all grid points are time series with zero means and standard deviations of 1) to ensure minimal influence from magnitude differences among the grid points.

Following this, the next step is to calculate the correlation matrix. A few issues arise when computing the correlation matrix. The first is that the initial dataset consists of 1,470,150 grid points (9801 data points by 150 levels in the NARR dataset) for 128 TO events where the 150 levels are the 29 levels of the NARR plus the surface, times the 5 meteorological variables (temperature, pressure, humidity, u and v wind components), and it is important to determine along which dimension **R** should be calculated, with **R** being the correlation matrix. The next question is whether we are interested in the correlations between the rows (grid points) or the columns (events)? Richman (1986) describes these scenarios as the *S* mode and the *T* mode. Since the relationship between the TOs is of interest in this study (and was more computationally tractable), the *T* mode analysis was chosen to calculate the correlation matrix.

The correlation matrix, **R**, is put into an eigenvalue matrix, D, with an associated eigenvector matrix, V, given by:

$$\boldsymbol{R} = VDVT \tag{3}$$

Usually an eigensolver (R, Bell Laboratories 2011), can be used to solve this equation and the resulting eigenvector matrix defines a new coordinate system. The resulting eigenvalue-variance relationship suggests that a subset of the original eigenvalues that are associated with the eigenvector matrix V, describes most of the meaningful variability in Z. To prevent fitting the composites that represent noise, or those with lower variability eigenvalues, the matrix V is to be truncated by a scree test prior to the final computation of F to keep the eigenvalues that contain the highest variability. The scree test looks at a plot of the eigenvalues and once the data has leveled off, the remaining eigenvalues are rejected, and the first portion of the data is retained. Figure 3 shows the resulting scree test used to select the number of clusters to include for the cluster analysis. While this is a subjective test, it is a simple method to reduce the data to ensure the important data is kept.

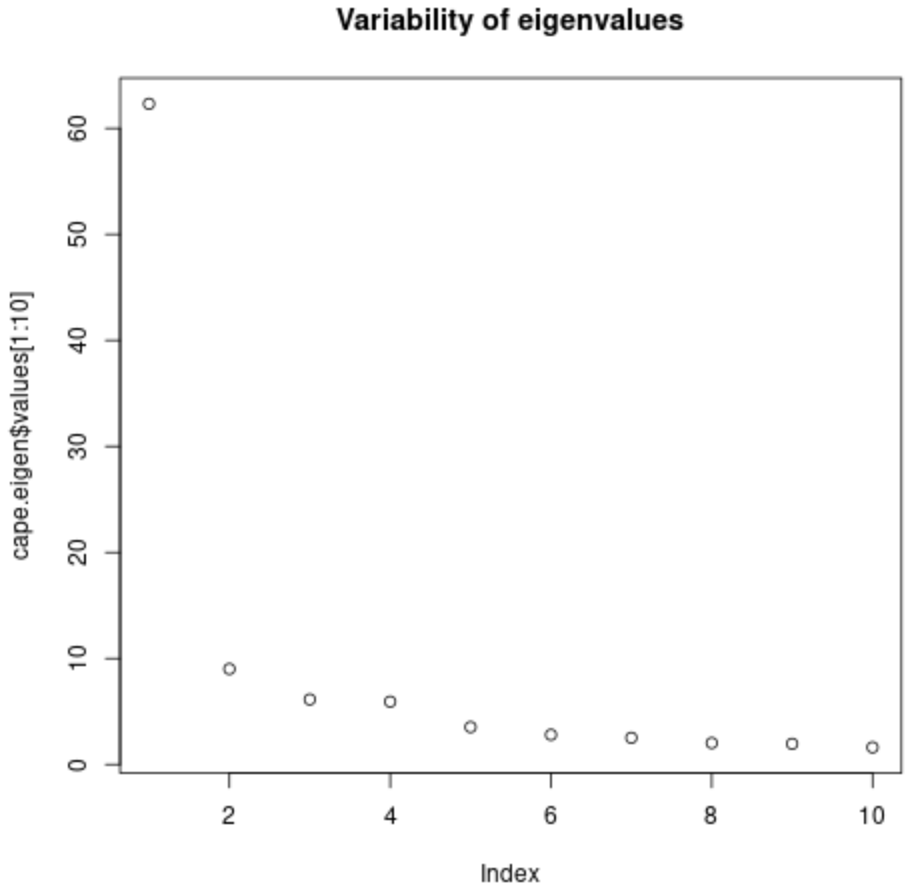

Figure 3 Scree plot used to identify 4 clusters used in cluster analysis.

To obtain the correct number of eigenvalues to keep for this study, the truncated **V** and **D** matrices are used to calculate the loading matrix, **A**, which is defined as $\mathbf{A} = \mathbf{V}\mathbf{D}^{1/2}$. The loading matrix will have a dimensionality of 128 x the number of retained PCs. These loadings were input into a nonhierarchical k-means cluster analysis (Wilks 2011) of the PC loadings for each TO. This is beneficial because the PC loadings by themselves isolate some of the localized significant features that do not fully represent all layers of the atmosphere. When the localized features are combined through clustering, a full picture of the atmospheric profile can be

obtained. Using a k-means cluster analysis means the number of composite map types for each outbreak must be known before conducting the analysis. The composite maps will be obtained through averaging as to ensure the data is not skewed in any way. To find how many clusters were to be used in the cluster analysis, a dendrogram was used to determine the appropriate cutoff for each cluster. The cutoff location resulted in four clusters being utilized for the cluster analysis. This location was chosen because it appeared there were four distinct clusters of data within the dendrogram that would provide the best results while having the greatest distinction between the clusters. Figure 4 shows the dendrogram used to determine the number of clusters to use in the cluster analysis, with the red line across the graph showing where the line was drawn to designate the four separate clusters.

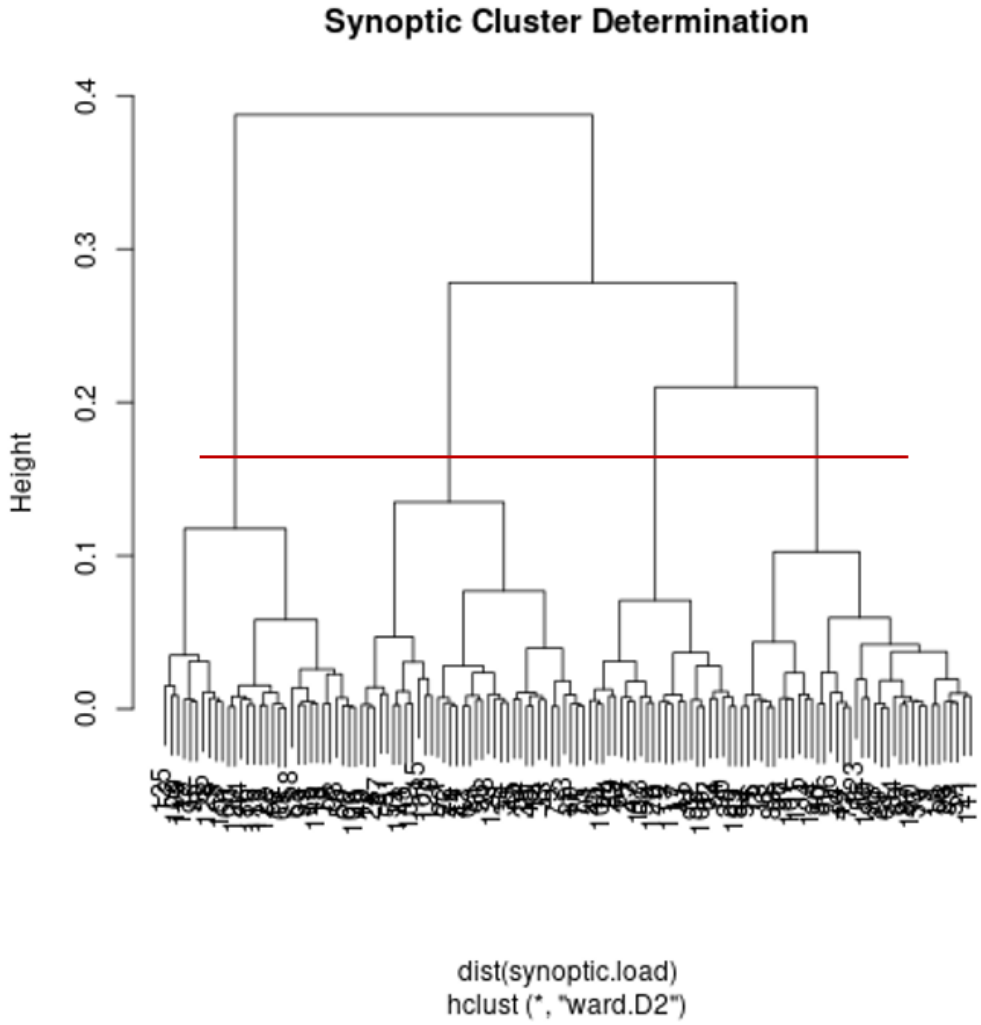

Figure 4 Dendrogram used to determine the number of clusters to be used in the cluster analysis. Red line represents the line used to separate the four clusters.

From the cluster analysis and the PCA, composite map types will be able to be plotted to analyze synoptic and mesoscale severe weather parameters that characterize a TO. From these map types, parameters such as CAPE, CIN, EBS, SRH, and EHI among others will be computed to portray the spatial structures of these variables underlying the different TO map types. These fields, when combined with detailed synoptic scale analyses from the raw NARR composite

fields, will provide a synoptic and mesoscale depiction of events meeting this new TO criteria. Finally, these patterns will be compared with past work to determine new spatial patterns revealed by this updated approach to defining a TO.

CHAPTER IV

RESULTS

The cluster analysis of the TOs produced four distinct clusters, meaning four separate types of TOs, each with its own set of characteristics. The clusters have their own individual values regarding synoptic and mesoscale patterns as well as spatial and temporal patterns. From each cluster, synoptic scale, mesoscale, thermodynamic and kinematic features, as well as composite indices, were plotted to illustrate the intensity and colocation of the different variables related to the associated TO.

Climatological Characteristics

After the cluster analysis was completed, new patterns emerged within the clusters showing changes in TO characteristics over time. Notably, TOs in this dataset became more common east of the Great Plains over time, as did the synoptic scale and mesoscale setups for those TOs. This geographic shift led to some interesting differences in the synoptic environments that are discussed below.

To gain information of the monthly timing of TOs within the clusters, the Julian day for each TO was calculated, as was the median Julian day for each cluster. This result showed that cluster 1's TOs has a median date of May 11, while cluster 2 has median of May 16. Cluster 3 has a median of April 26, and cluster 4 has a median of April 22. The values in Figure 5 show the TOs typically occurred within the same 4-month period (March-June), with outliers showing

in January and November. The values within each histogram in Figure 5 also show a range of Julian dates in which the TOs occur within. Typically, it is across a span of ~100 days, showing the main period in which tornadoes, and subsequent TOs, are most common in the United States. Cluster 3 shows a secondary TO season during the cool season (late fall through winter) when the remaining warm, moist air is colliding with the cold fronts moving in from the northern United States. Cluster 1,2, and 4 do not show this secondary season, indicating that the TOs within these clusters are seen during the peak months for TOs (March-June). However, from the histograms it is evident that TOs in clusters 1 and 2 typically occur slightly later than the TOs in cluster 3 and 4, in correlation with the median Julian day values.

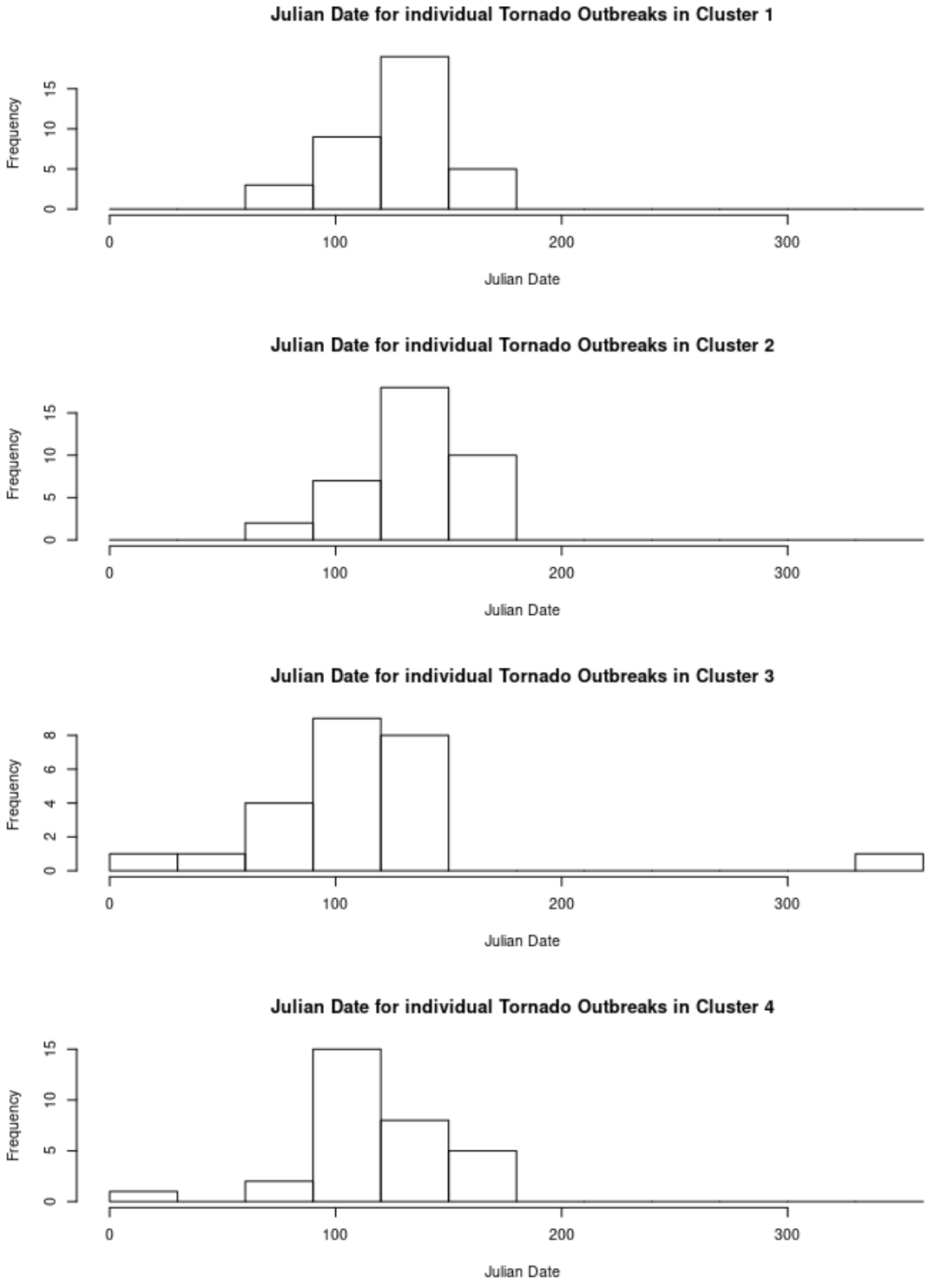

Figure 5 Histogram for each cluster showing Julian date of individual TOs throughout the year. Each bin represents 30 days.

Another noticeable change between the clusters is the average year of the TOs within that cluster. This was found by taking the mean and median of the years within each cluster, ending up with an average year for each cluster. Cluster 1 saw a mean and median outbreak year of 1992, with 80% of its outbreaks occurring prior to 2000. Cluster 2 has an average year of 1997 with a median year of 2000. Its outbreaks are mainly split between the 1990s and the early 2000s. Cluster 3 has an average year of 1998 with 79% of its outbreaks coming from the 1990s and 2000s. Cluster 4 has an average outbreak year of 2001 and a median of 2002. Around 33% of the outbreaks within cluster 4 have occurred over the last 6 years of the study period (2006-2011). Figure 6 shows the distribution of the TOs in each cluster based off the year they occurred in. While there is not much difference between clusters 2 and 3, there is a difference between clusters 1 and 4. The graph is able to back up the numbers provided and show that cluster 1's TOs occurred earlier in the record when compared to those in cluster 4.

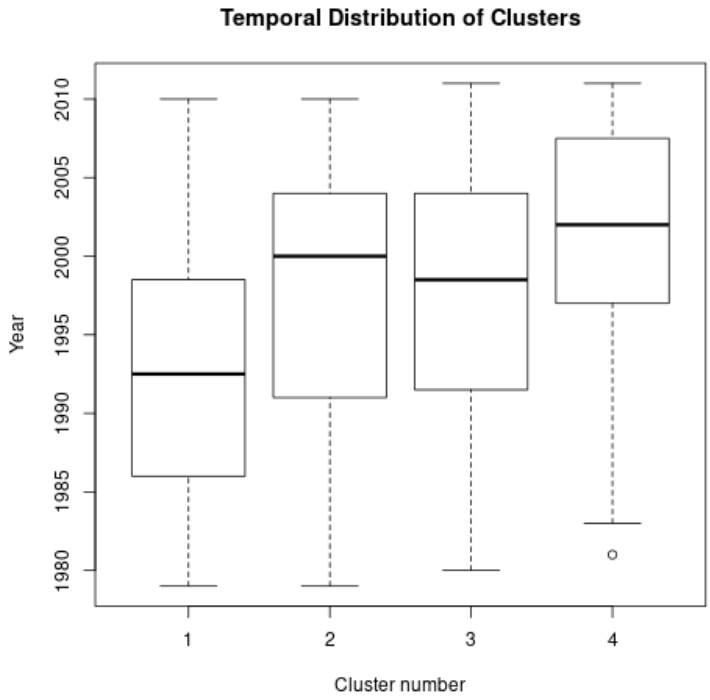

Figure 6 Boxplot showing the distribution of TOs in each cluster as related to the year in which they occurred.

Another distinguishable factor between the clusters is the change in the average location of a TO within each cluster. This was found by taking the mean of the latitude and longitude, separately, for each TO within each cluster. The latitude and longitude for each TO represent the center point of the TO itself. Figure 7 shows the average TO location for each cluster. The numbers on the map correspond to the cluster each location is representative of. It was found that the average location of TO in cluster 1 occurred around Union City, OK, which is WSW of Oklahoma City, OK. The average location of a TO in cluster 2 is in the SE corner of Nebraska. For cluster 3, the average location of a TO was in the NE portion of Arkansas. In cluster 4, the average location of a TO was found to be just east of Louisville, KY. From this it is evident that

there is a change in spatial distribution of the TOs between the 4 clusters. While there are outliers in each cluster, a majority of the TOs do occur in or around the average location for each cluster.

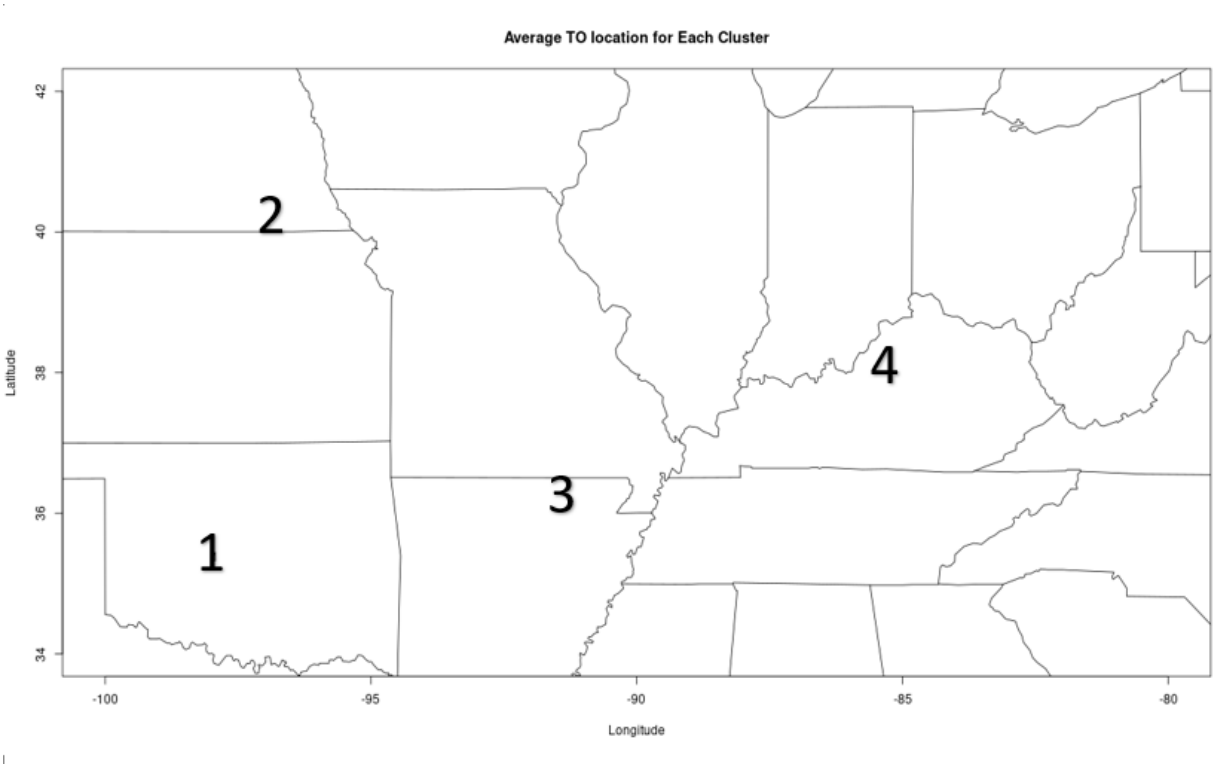

Figure 7 Map illustrating the average location of a TO in each cluster as designated by the number on the map. 1 represents cluster 1, 2 represents cluster 2, 3 represents cluster 3, 4 represents cluster 4.

Figure 6 and Figure 7 help illustrate an overall larger pattern and trend that is occurring. Figure 6 provides evidence that the TOs are beginning to resemble the characteristics of cluster 4 more often than that of the other 3 clusters. Figure 7 shows that the average location of a TO in each cluster differs with the average locations being spread across the Great Plains in the first 2 clusters and moving into the Southeast and Ohio River Valley for clusters 3 and 4. It is reasonable to infer that not only are the TOs becoming more like those in cluster 4 as time

progresses, but also that the TOs are moving further to the east as well. The average location for each cluster has progressively shifted eastward as you move through the dataset in chronological order. A study by Gensini and Brooks (2018) backs up this claim as they found tornado frequencies increasing in the areas to the east of the Great Plains (Texas, Oklahoma, Kansas, Nebraska, and the Dakotas) and tornado frequency decreasing over the Great Plains. Gensini and Brooks (2018) study uses data from 1979-2017 while this study only uses 1979-2011. The difference is only 6 years, but the trend is already evident once the dataset ends in 2011.

Synoptic and Mesoscale Composite Results

Several important plots for each cluster were created to help illustrate the mesoscale and synoptic scale environments. This includes various pressure level maps as well as thermodynamic parameters and composite indices mentioned previously in the study. For simplicity, the TO region was set across portions of Missouri and Illinois and are not geo-referenced. The map is provided for a sense of scale since each TO is georeferenced to the same position in the grid. Additionally, an example case from each cluster was provided to demonstrate the representativeness of the cluster patterns derived from the analysis.

Cluster 1

Synoptic scale features within cluster 1 show many of the expected signs of a TO including the positioning of the jet streak (Rose et al. 2004), the upper-level trough west of the TO region (Mercer et al. 2011), and the collocation of the LLJ with the TO region (Weaver et al. 2012). Figure 8a shows the location of the jet stream over the TO area. The TO region is located within the exit region of the upper-level jet streak and is also on the right side of the exit region. It also shows the location of the upper-level trough that is west of the TO area, signifying a

pattern in which TOs tend to occur according to Mercer et al. (2011). This pattern also appears in the 500mb vorticity map in Figure 8b. This figure also shows CVA leading into the TO region, which helps provide the proper rising motion associated with the warm sector of an extratropical cyclone. Moving into the lower portions of the atmosphere you can see that the LLJ in Figure 8c is co-located over the TO region as well. The added boost of the LLJ will provide wind shear in the lower portions of the atmosphere. Figure 8d shows WAA and moisture being transported into the TO region, providing the warm and moist air needed to destabilize the atmosphere. By the orientation of the specific humidity values in Figure 8d it appears as if these events are affected by a near surface dryline. Drylines are common sources of surface convergence and uplift needed for thunderstorm generation in the southern Great Plains.

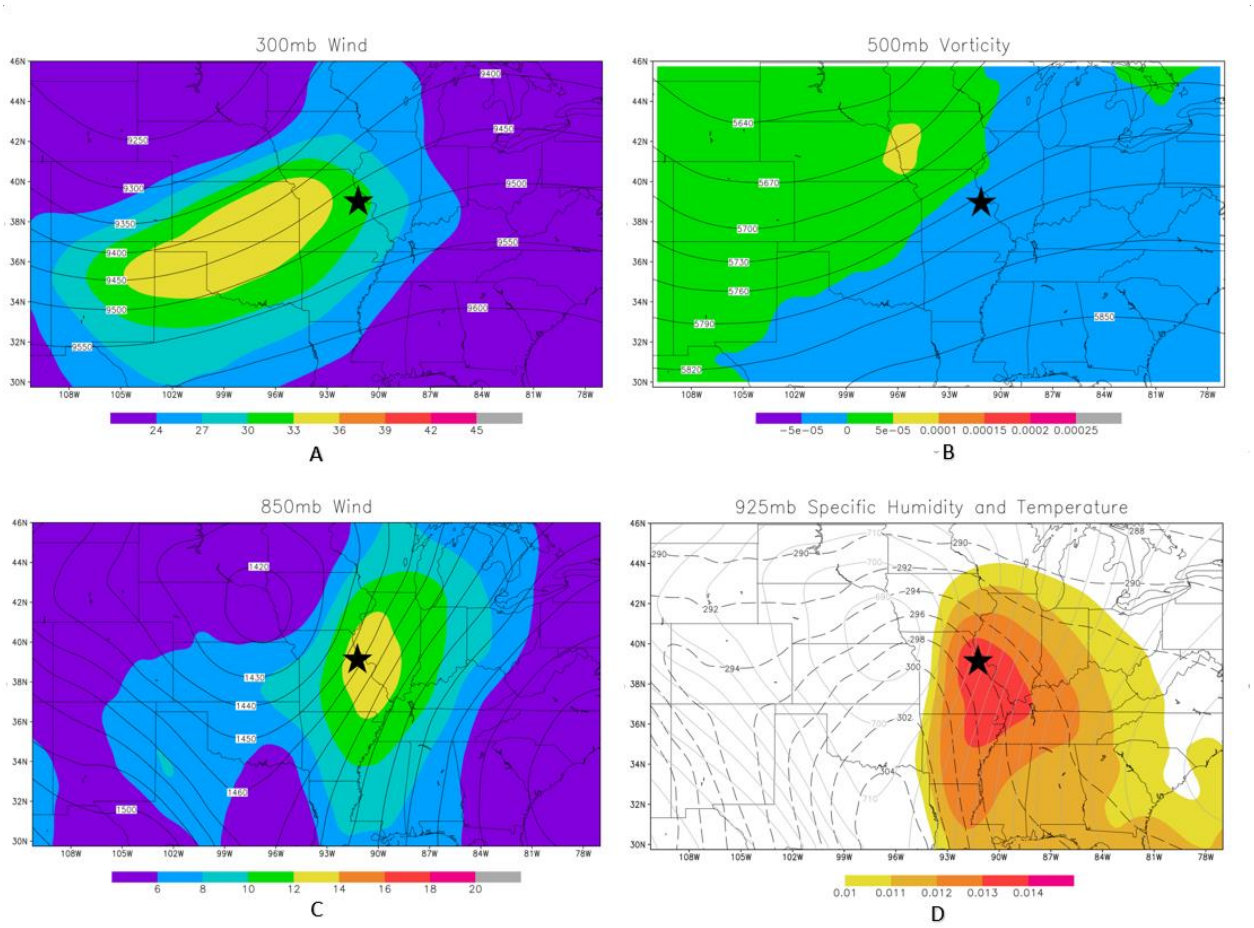

Figure 8 Composite maps from cluster 1. 300mb wind speed (m/s) (shaded) and geopotential height (m) (A), 500mb vorticity (s^{-2}) (shaded), and geopotential height (m) (B), 850mb wind speed (m/s) (shaded) and geopotential height (m) (C), 925mb Specific humidity (kg/kg) (shaded), temperature (K), and geopotential height (m) (D). The black star represents the center of the TO.

The mesoscale features in cluster 1 align with those previously mentioned in the introduction of this study. CAPE values reach higher than 3000 J/kg (not shown) over the TO region while CIN values in the same location are very low and are more than -20 J/kg, meaning values are less than -20 J/kg. LCL values are between 900mb and 850mb (not shown), which is a range of about 950 meters to 1450 meters for a given TO. These are roughly consistent with Rasmussen and Blanchard (1998) and any differences were attributed to the composite averaging

smoothing out higher values of the mesoscale fields. Rasmussen (2003(b)) found 0-1km SRH to be a distinguishing factor between tornadic supercells and non-tornadic supercells and it showed to be a changing feature across each cluster. For cluster 1, Figure 9b shows 0-1km SRH values of 110 m^2/s^2 and higher within the TO region. While these values are not overly impressive or very high, they are enough to help sustain an updraft within a mesocyclone. Similar to the LCL values, the higher values of 0-1km SRH are diminished by the averaging of the composites.

EBS is another feature that appears to change across each cluster and is illustrated in Figure 9a. It is important because it normalizes the shear values for both shallow and taller storms, which allows for a more realistic assessment of a storm's vertical profile (Thompson et al. 2007). According to NOAA's SPC, supercells become more probable as the EBS increases beyond 25-40 knots. For cluster 1, there was a wide area with values over 15-24 m/s, which equates to 29-46 knots. This indicates a wide area across the TO region had the enhanced potential for supercells to develop. Another feature is the 0-1km EHI, noted in Figure 9d. For this cluster, the highest valued contour is 2.5, which exceeds the threshold of 2 that was found by Rasmussen and Blanchard (1998) and indicates a larger probability of supercells. Other composite indices such as the SCP and STP back up the other severe weather parameters in showing that the TO region is an area where supercells, and even tornadoes, are likely to develop. SCP values exceed 10, which passes the threshold of 1 determined by Thompson et al. (2003(b)) suggesting a higher likelihood of supercells occurring. The STP value in Figure 9c reaches over the threshold of 1 as well at a value of 1.2, indicating it is reasonable to distinguish between tornadic supercells and non-tornadic supercells based upon the maximization of forecast skill at a value of 1 (Thompson et al. 2003(b)). Finally, as noted previously, each TO in the data set had an associated ranking index from Shafer and Doswell (2010) that describes the outbreak

strength. TOs in cluster 1 had a median value of 0.903, which is on the high end of the ranking index distribution (**Error! Reference source not found.**).

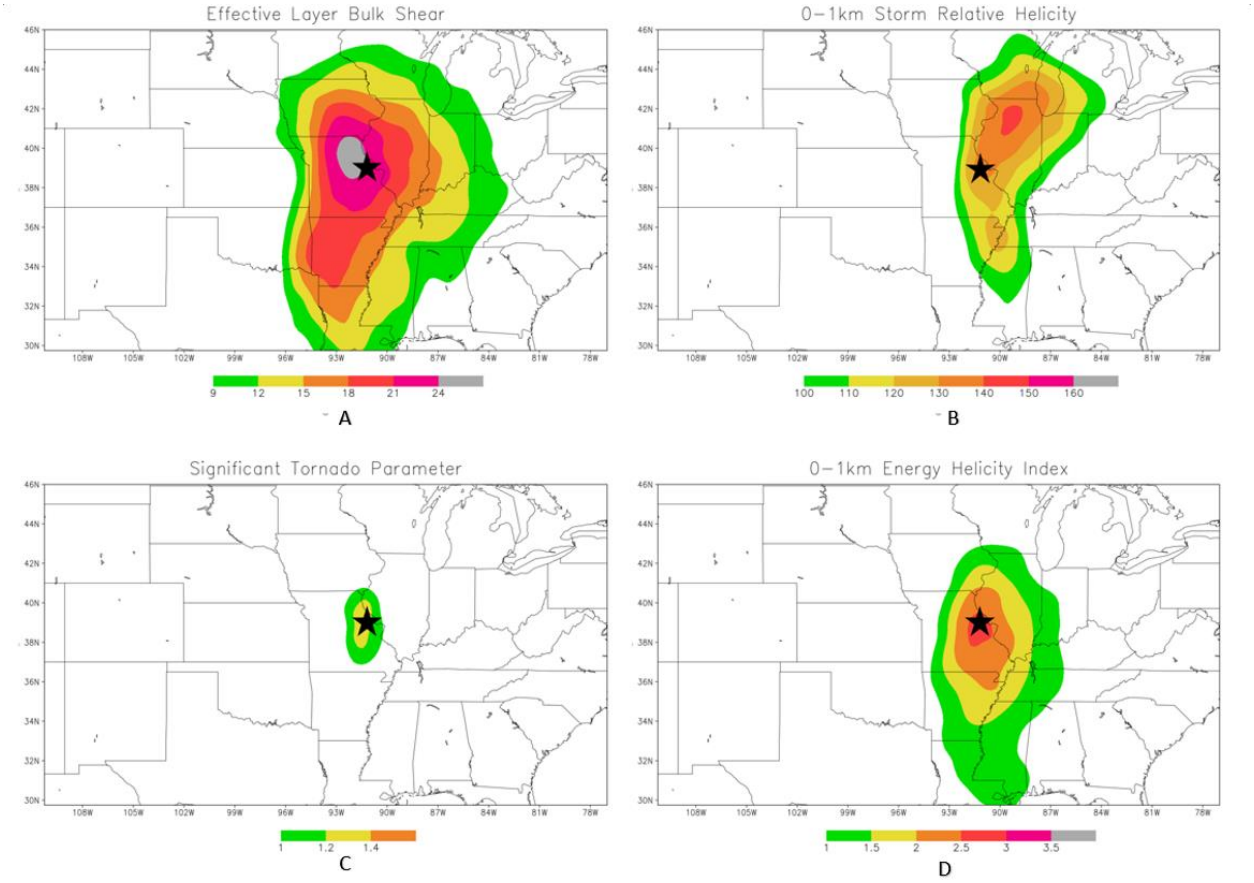

Figure 9 Composite maps from cluster 1. Effective layer bulk shear (m/s) (A), 0-1km Storm relative helicity (m^2/s^2) (B), Significant tornado parameter (C), 0-1km Energy helicity index (D). The black star represents the center of the TO.

To ensure the composites were representative of their constituent cases, one real-world example from the cluster 1 case set was selected (April 22, 2010) for analysis. The 300mb map in Figure 10a shows general TO region is in the right exit region of the jet streak as it is in Figure 8a but the jet streak within the example is 10 to 15 m/s stronger. Figure 10b shows CVA

occurring around the outbreak region as it does in the composites, as well as the upper-level trough being located to the west of the TO region (Mercer et al. 2011). The LLJ is also present in Figure 10c with similar values to those in the composites. In Figure 10d the moisture tongue is present and the vertical orientation of it suggests a dryline event occurred to aid in the development of the TO. The magnitudes of the specific humidity available is similar when compared to the composite. Surface CAPE values in the example are closer to 2000 J/kg (not shown) across the TO region while in the composite values are between 3000 J/kg and 3500 J/kg. LCL values are also like the composite with LCL values being close to 850mb (not shown), or roughly 1450 meters. Overall, this example matches up well with the composite and provides a real-life case study as to the types of TOs associated with cluster 1.

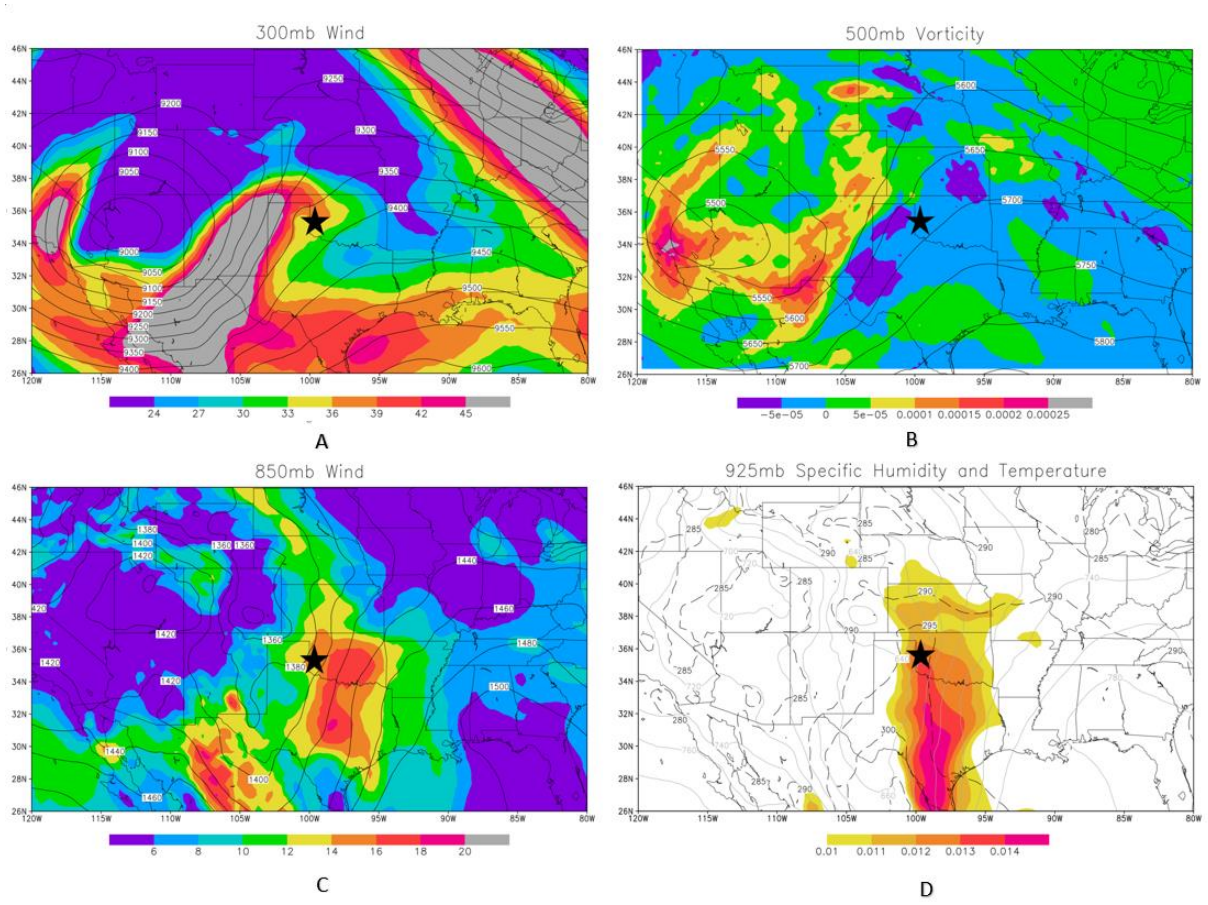

Figure 10 Maps from example outbreak from April 22, 2010. 300mb wind speed (m/s) (shaded) and geopotential height (m) (A), 500mb vorticity (s^{-2}) (shaded) and geopotential height (m) (B), 850mb wind speed (m/s) (shaded) and geopotential height (m) (C), 925mb specific humidity (kg/kg) (shaded), temperature (K), and geopotential height (m) (D). The black star represents the center of the TO.

Cluster 2

Cluster 2 contains TOs that occur throughout the Great Plains, reaching from the Dakotas down into Oklahoma. This means that some of the same characteristics that were present in cluster 1 are also present within cluster 2. The TO region is in the right exit region of the upper-level jet streak as noted in Figure 11a. The jet streak in cluster 2 is stronger by 3-6 m/s than the one found in cluster 1 and is also located slightly further to the SW in relation to the TO region.

So, there is not much difference in the strength of the jet streak between the first two clusters. There is CVA occurring across the TO region, as seen in Figure 11b, with the upper-level trough located to the NW of the outbreak area. Within the lower levels at 850mb in Figure 11c, the LLJ is stronger in this cluster than to cluster 1 by up to 4 m/s around the TO region. Coupling this along with the 925mb humidity values in Figure 11d, the LLJ is co-located with the stronger moisture tongue (0.001 kg/kg stronger). WAA is also shown to be present in Figure 11d, meaning the warm and moist air is advected into the TO region ahead of the dryline or cold front. The stronger LLJ and increased moisture content signifies a stronger TO environment because of the greater wind shear as well as the higher moisture available. When more moisture is available, it becomes easier to lower the LCL which can be beneficial to storms capable of producing tornadoes.

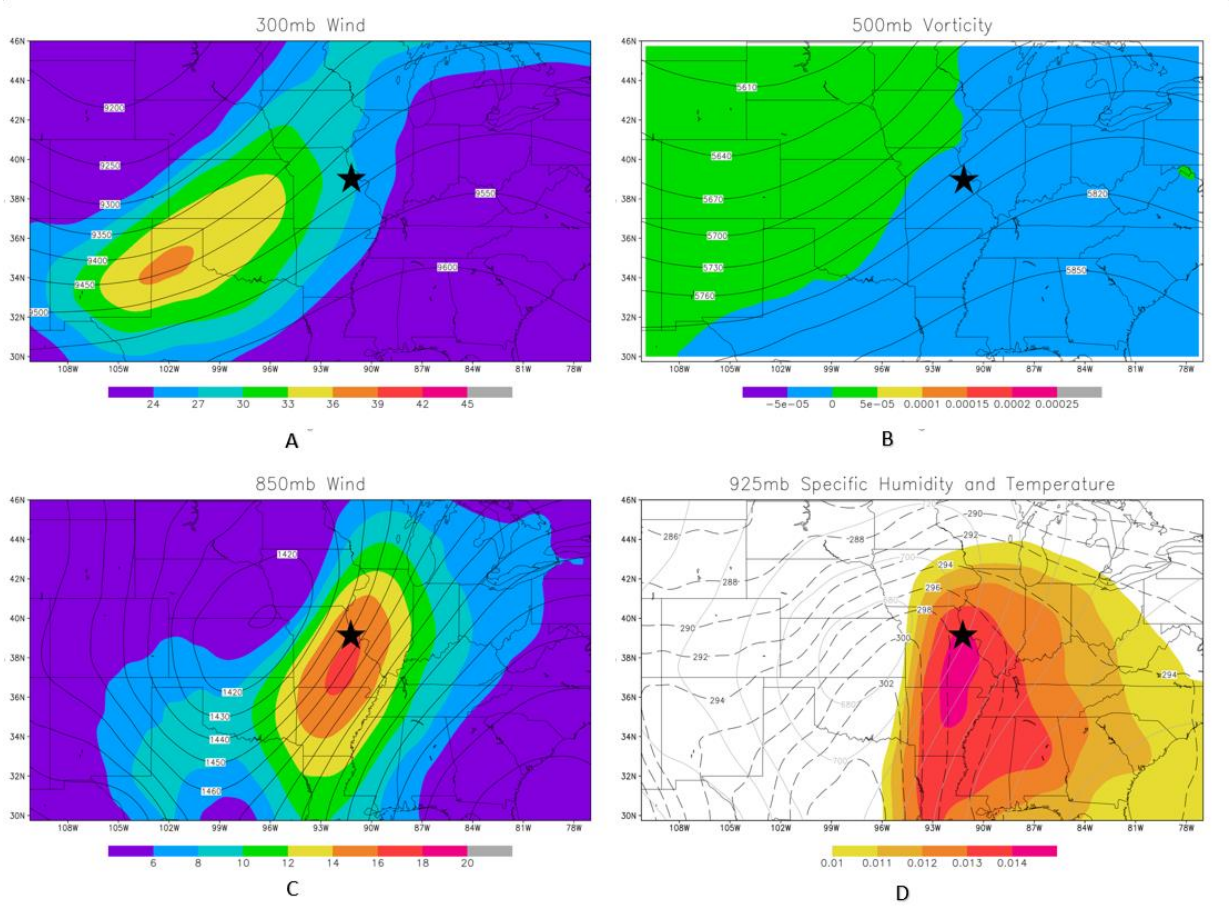

Figure 11 Composite maps from cluster 2. 300mb wind speed (m/s) (shaded) and geopotential height (m) (A), 500mb vorticity (s^{-2}) (shaded) and geopotential height (m) (B), 850mb wind speed (m/s) (shaded) and geopotential height (m) (C), 925mb specific humidity (kg/kg) (shaded), temperature (K) (black dashed), and geopotential height (m) (D). The black star represents the center of the TO.

The mesoscale features that are present in cluster 1 also appear in cluster 2, except the values for some of them are more extreme. CAPE/CIN values did not change too much from before with the CAPE values over the TO region being above 3500 J/kg (not shown) whereas the other cluster contained those between 3000 J/kg and 3500 J/kg. CIN values are slightly lower with values of -20 J/kg to -30 J/kg. The average LCL within cluster 2 still appears to offer a value between 850mb and 900mb (not shown), which is about 950 meters to 1450 meters. 0-1km

SRH values in Figure 12b shows values over the TO region of 140 m^2/s^2 or greater, which is enough to help sustain an updraft within a mesocyclone. EBS values in Figure 12a range from 18-22 m/s across the TO region, which is a range of 35-42 knots. This is enough to conclude supercells will be likely to form within the TO region (Thompson et al. 2007). Another value that can also confirm this is the SCP (not shown), which has a value well over the threshold of 1 mentioned by Thompson et al. (2003(b)). For cluster 2 the peak value within the TO region is 13 and across the area reaches up to 9. STP values in Figure 12c also goes above the threshold of 1 and reaches a peak of 1.4, meaning it is reasonable to distinguish between significant tornadic supercells and non-tornadic supercells (Thompson et al. 2003(b)). In Figure 12d, 0-1km EHI values show they are over the threshold of 3 defined by Rasmussen and Blanchard (1998), which says that it indicates the enhanced possibility of a strong mesocyclone-induced tornado to occur with a strength of F/EF2 or higher. The ranking index for cluster 2 gave a median value of 1.202 for each TO, which indicates that the TOs in cluster 2 were slightly stronger than those within cluster 1. Given the somewhat similar characteristics and patterns between the two clusters, cluster 2 is a stronger version of cluster 1.

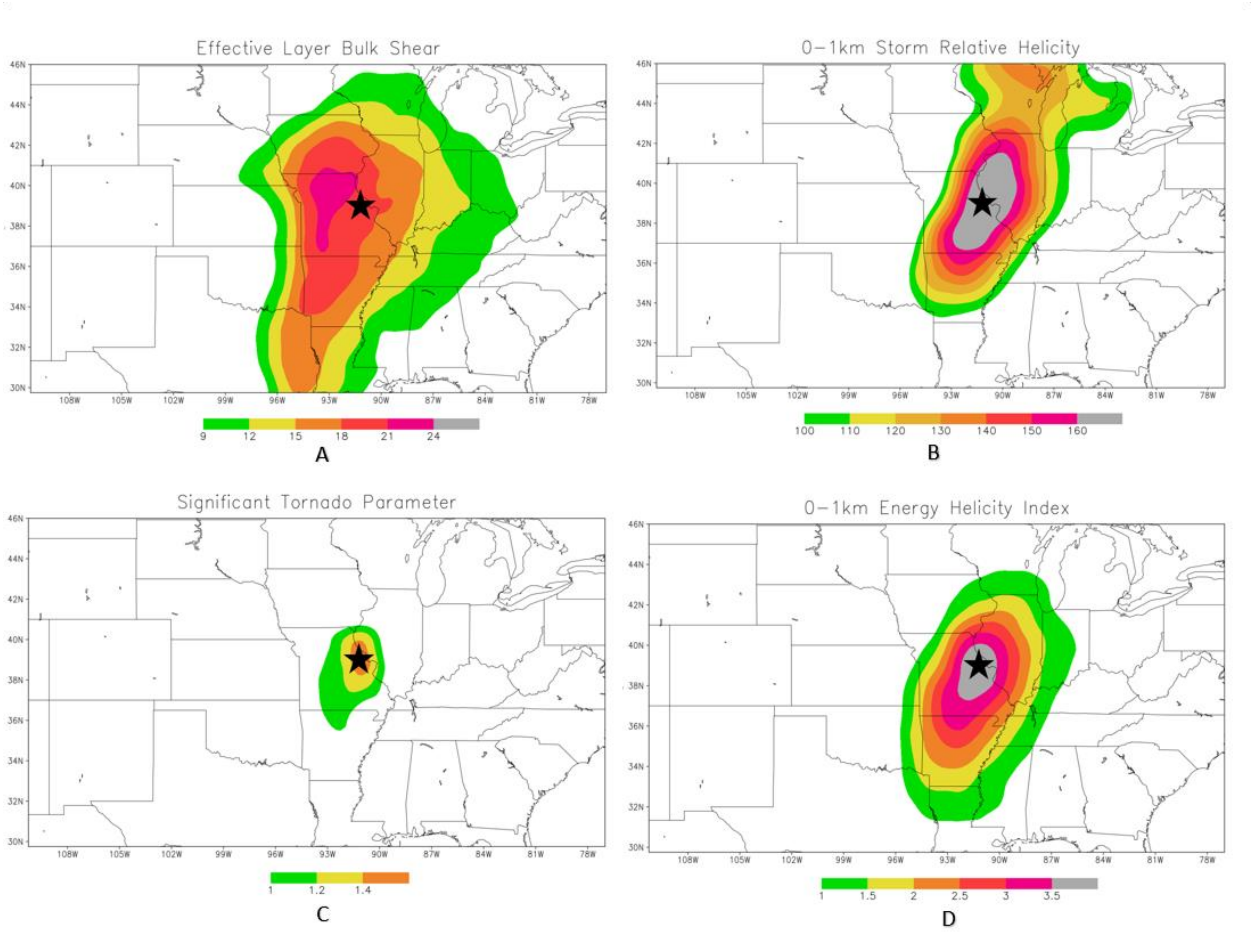

Figure 12 Composite maps from cluster 2. Effective layer bulk shear (m/s) (A), 0-1km storm relative helicity (m^2/s^2) (B), significant tornado parameter (C), 0-1km energy helicity index (D). The black star represents the center of the TO.

A representative example of a TO that occurred in cluster 2 happened on June 24, 2003, spanning across parts of Minnesota and Nebraska, with the large portion of the tornadoes occurring around Sioux Falls, South Dakota. This TO included 2 F4, 2 F3, and 9 F2 tornadoes, along with many more F1 and F0 tornadoes. The upper-level jet streak from the example in Figure 13a contains the TO region in its right exit region, the same as the composite map from Figure 11a. The only difference is that the jet streak from the June 24, 2003 TO is more than 6 m/s stronger than the composite. CVA is still evident in Figure 13b within the example TO and is

receiving slightly more CVA when compared to the composite with up to $1x10^{-4}$ s^{-2} being advected into the TO region where the composite only sees $5x10^{-5}$ s^{-2}. The upper-level trough is also located to the west of the TO region just as in the composite. The LLJ in Figure 13c also appears to be stronger when compared to the composite with values exceeding 20 m/s near the TO region whereas the composite only reaches up to 18 m/s. Once again there is also an increase in the humidity values in Figure 13d as well with a large area having values higher than 0.014 kg/kg with the composite only having a small area around the TO region reaching this value. Surface based CAPE values are also similar between the composite and the example. The composite has values up to 3500 J/kg-4000 J/kg while the example has values between 4000 J/kg and 4500 J/kg (not shown). LCL heights are the same with both occurring between 900mb and 850mb (not shown). The higher CAPE values coupled with the stronger LLJ and higher CVA make this environment more suitable for a TO than the composite environment. This example fits the composite well and shows the same overall patterns and characteristics as the composites for cluster 2.

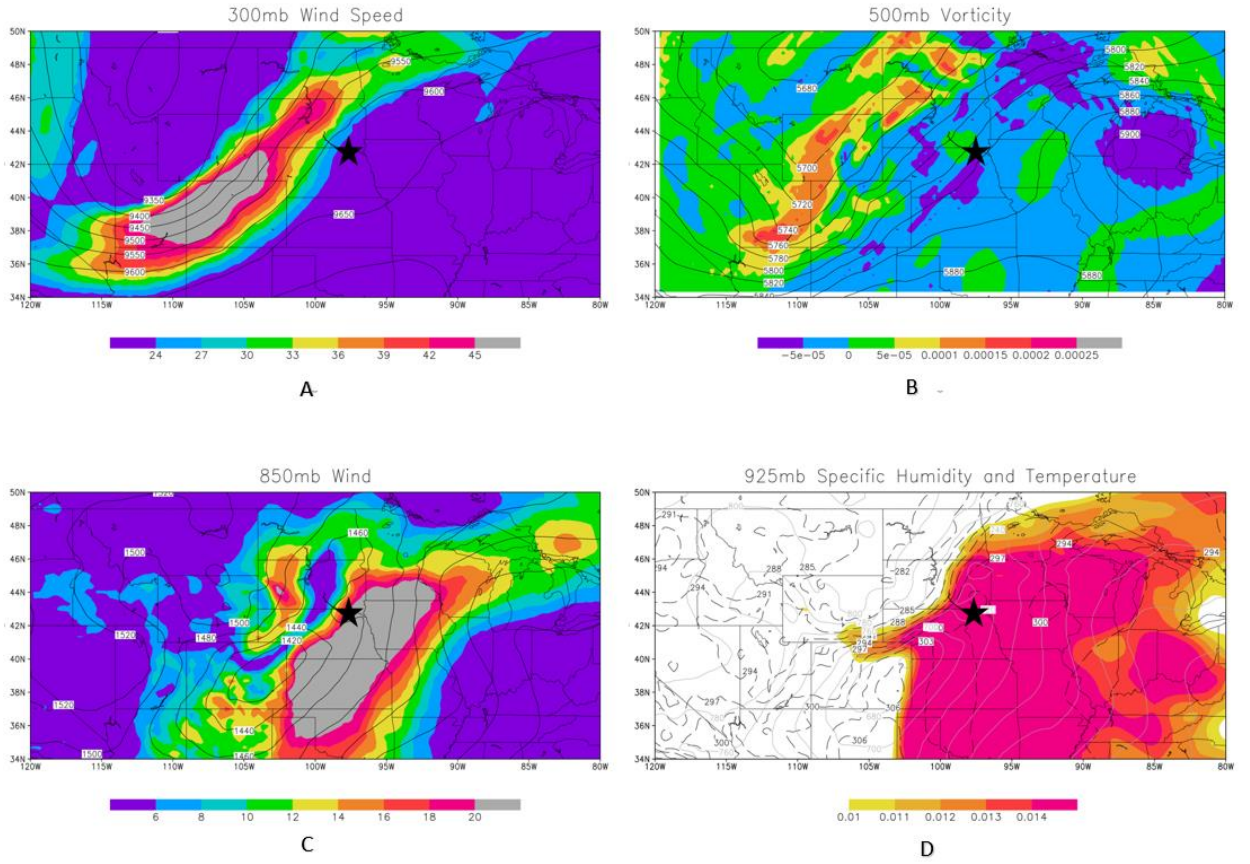

Figure 13 Maps from example outbreak from June 24, 2003. 300mb wind speed (m/s) (shaded) and geopotential height (m) (A), 500mb vorticity (s^{-2}) (shaded) and geopotential height (m) (B), 850mb wind speed (m/s) (shaded) and geopotential height (m) (C), 925mb specific humidity (kg/kg) (shaded), temperature (K) (black dashed), and geopotential height (m) (D). The black star represents the center of the TO.

Cluster 3

The synoptic scale patterns found within cluster 3 are like those from the first two clusters, except the features are more intense. The upper-level jet streak in Figure 14a is stronger and is located just to the SW of the TO region, which is in the right exit region of the jet streak. The jet streak contained a maximum wind speed of ~42 m/s, which is higher relative to clusters 1

and 2. The 500mb map in Figure 14b shows that there is CVA occurring over the TO region and is roughly the same magnitude as the CVA in clusters 1 and 2. This negatively tilted trough coupled with the other stronger synoptic scale dynamics suggest this cluster is likely strongly synoptically forced. The LLJ is present over the TO area in Figure 14c and is similar in strength to the LLJ in cluster 2. 925mb values in Figure 14d are not as high in this cluster compared to cluster 2 and do not exceed 0.014 kg/kg. The same map also shows that WAA is occurring across the TO region.

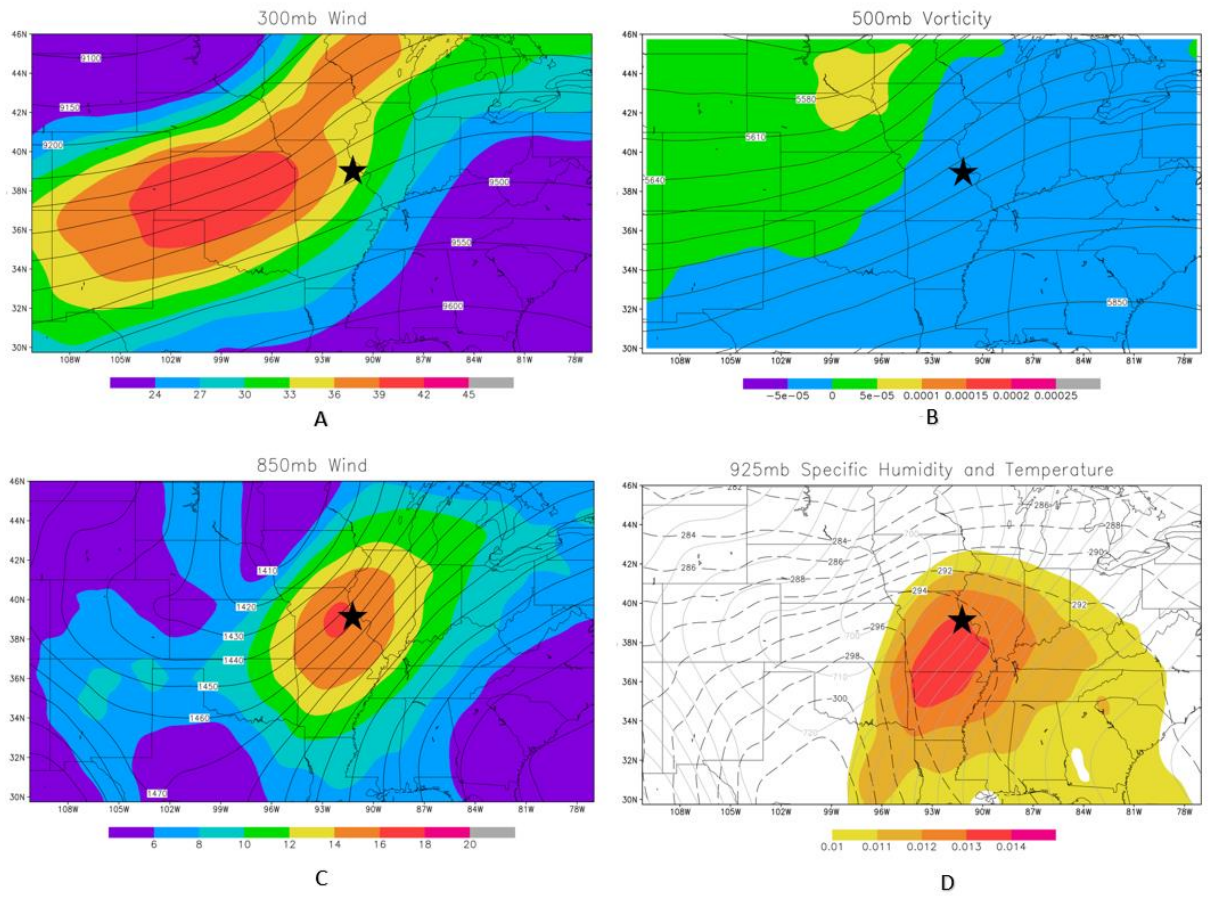

Figure 14 Composite maps from cluster 3. 300mb wind speed (m/s) (shaded) and geopotential height (m) (A), 500mb vorticity (s^{-2}) (shaded) and geopotential height (m) (B), 850mb wind speed (m/s) (shaded) and geopotential height (m) (C), 925mb specific humidity (kg/kg) (shaded), temperature (K) (black dashed), and geopotential height (m) (D). The black star represents the center of the TO.

CAPE values from cluster 3 are between 2500 J/kg and 3500 J/kg (not shown) across the TO region with CIN values at or close to 0 J/kg, so there is not a "cap" on the atmosphere within the composite. The lower CAPE values provide more evidence that the TOs within cluster 3 are heavily influenced by synoptic scale forcing. LCL values are around 900mb (not shown), or 950 meters across the TO region, which is lower than the previous two clusters. In Figure 15b, 0-1km

SRH values range from 140 m^2/s^2 to 160 m^2/s^2 across the TO region. The higher 0-1km SRH values provide further evidence that the TOs in this cluster are more synoptically forced and are not as thermodynamically strong as in clusters 1 and 2. Based on Figure 15a wind shear is a definitive factor in the development of TOs within cluster 3. EBS values show that the values reach above the threshold of 25-40 knots (40 – 47 knots in cluster 3) defined by Thompson et al. (2007) in which the probability of supercells forming increases. SCP values reach across the threshold of 1 and up to a value of 11 (not shown), showing supercells are likely to form across the area. STP values in Figure 15c also reaches across the threshold of 1 and reach a value of 1.4. The EHI index noted in Figure 15d meets the threshold of 2 across the TO region and reaches up to a 2.5, indicating a larger probability for supercells to develop according to Rasmussen and Blanchard (1998). Many of the mesoscale features once again point to the conclusion that supercell thunderstorms are very likely to develop given the state of the atmosphere over the TO region. The ranking index within this cluster had a median value of 1.613, which is higher than clusters 1 and 2. The ranking index helps assert the idea that this cluster is more synoptically force than clusters 1 and 2 because in order to obtain a stronger TO, more forcing from the synoptic scale is needed.

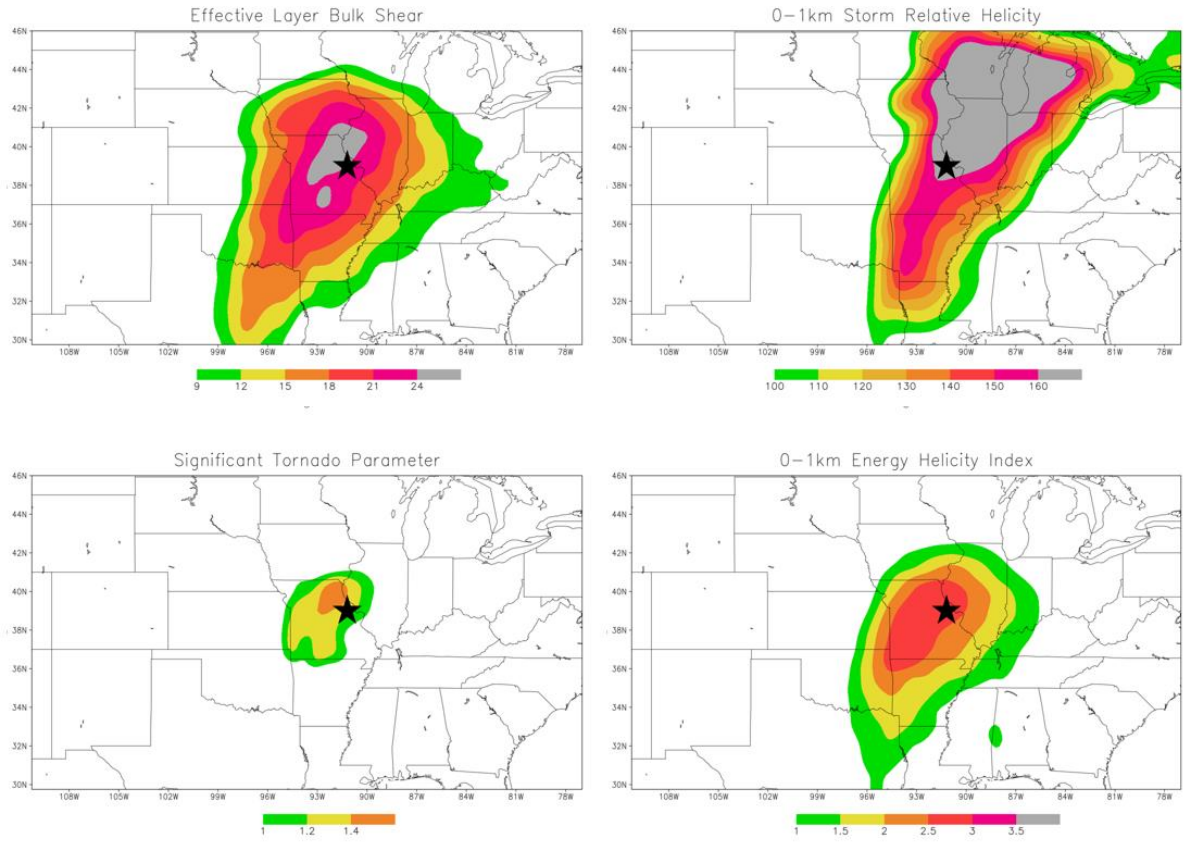

Figure 15 Composite maps from cluster 3. Effective layer bulk shear (m/s) (A), 0-1km storm relative helicity (m^2/s^2) (B), significant tornado parameter (C), 0-1km energy helicity index (D). The black star represents the center of the TO.

This example TO occurred on April 9, 2009 and provides a good example representative of cluster 3. The 300mb map in Figure 16a shows the TO region being in the exit region of the jet streak. The intensity of the jet streak is stronger compared to the composite map with values exceeding 45 m/s with the composite values being maxed out at 42 m/s. There is more CVA occurring in the example in Figure 16b compared to the composite and has less zonal flow as well, but the CVA is still occurring. At 850mb the LLJ in Figure 16c is stronger in the example than the composite map with values over 20 m/s compared to values up to 18 m/s in the

composite. The 925mb specific humidity values in Figure 16d are lower than those in the composite map but there is still WAA occurring into and ahead of the TO region. There is about 1000 J/kg less CAPE in the example compared to the composite. The example shows CAPE values of up to 2400 J/kg (not shown) whereas the composite is up to 3500 J/kg. The LCL values are roughly the same between the composite and the example with values between 950mb and 900mb. The weaker CAPE values and higher wind shear within the composites and the example TO imply once again that this cluster is being more synoptically forced compared to clusters 1 and 2. This evidence is also backed up by the higher-ranking index values associated with cluster 3.

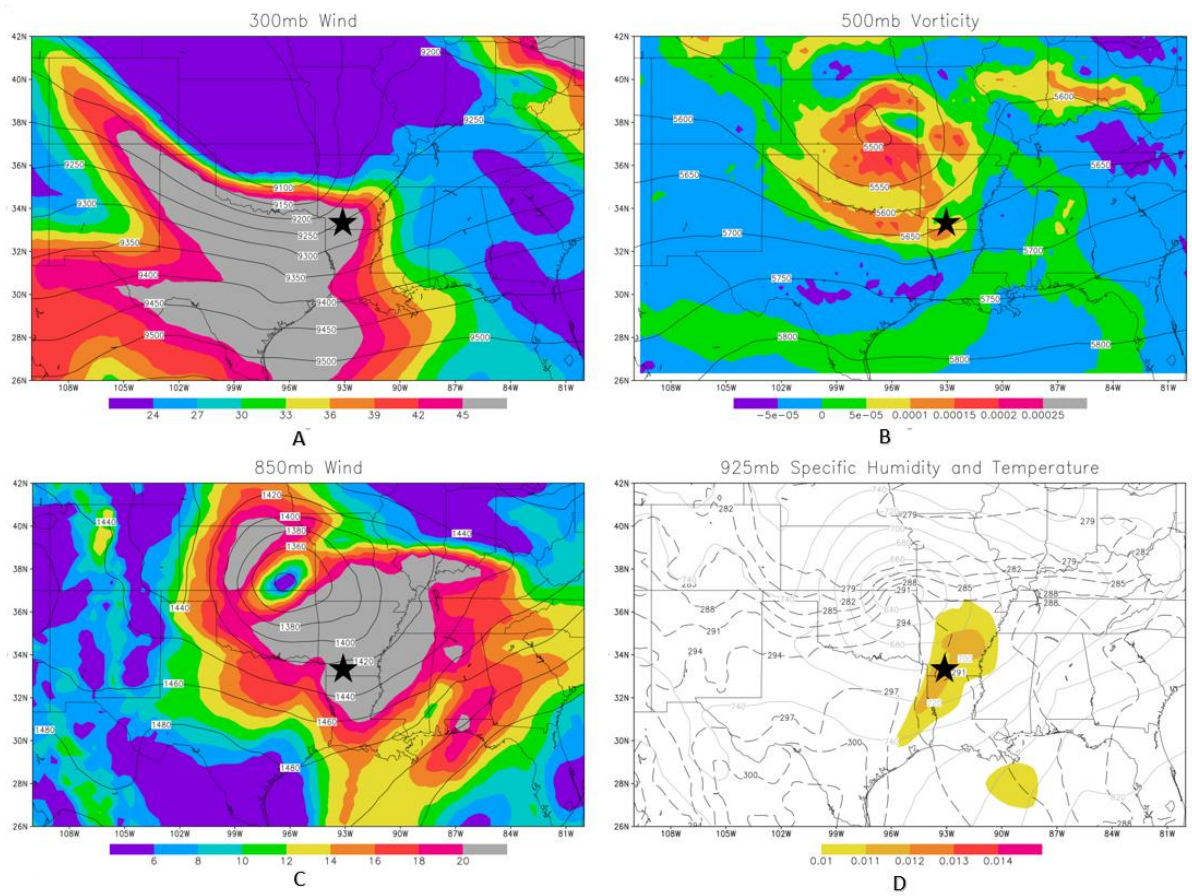

Figure 16 Maps from example outbreak from April 9, 2009. 300mb wind speed (m/s) (shaded) and geopotential height (m) (A), 500mb vorticity (s^{-2}) (shaded) and geopotential height (m) (B), 850mb wind speed (m/s) (shaded) and geopotential height (m) (C), 925mb specific humidity (kg/kg) (shaded), temperature (K) (black dashed), and geopotential height (m) (D). The black start represents the center of the TO.

Cluster 4

Cluster 4 contains TOs that occur east of the Great Plains and primarily those east of the Mississippi River since the average location of a TO is just east of Louisville, Kentucky. This cluster also has different synoptic and mesoscale characteristics than the other clusters, possibly due to the location of the TOs. In Figure 17a, the TO region is located towards the right exit region of the jet streak, the same as each of the previous clusters. The strength of the jet streak is

not as strong as in cluster 3 but is slightly stronger than the jet streak in cluster 2 due to wind speed values reaching up to 39 m/s. The 500mb vorticity map in Figure 17b shows very little CVA occurring, possibly because of the flow being more zonal than meridional at this level. The LLJ is co-located with the TO region and is weaker than those found in clusters 2 and 3 but stronger than the one in cluster 1. With these TOs being located further east and including TOs from the southeastern United States, one would expect the LLJ to be present at 850mb as depicted in Figure 17c. It is weaker than the LLJ found in clusters 2 and 3 and is roughly the same strength as the LLJ in cluster 1. The 925mb map in Figure 17d illustrates the placement of the moisture plume across the TO region. The specific humidity values are aligned where the gradient is oriented in the NW to SE direction. The overall pattern that is shown also appears to be more associated with the orientation of a cold front rather than that of a near-surface dryline. This alone indicates this cluster is being more synoptically driven than it is by the mesoscale environment. Temperature advection in cluster 4 is rather weak and there is not much WAA occurring within the composite.

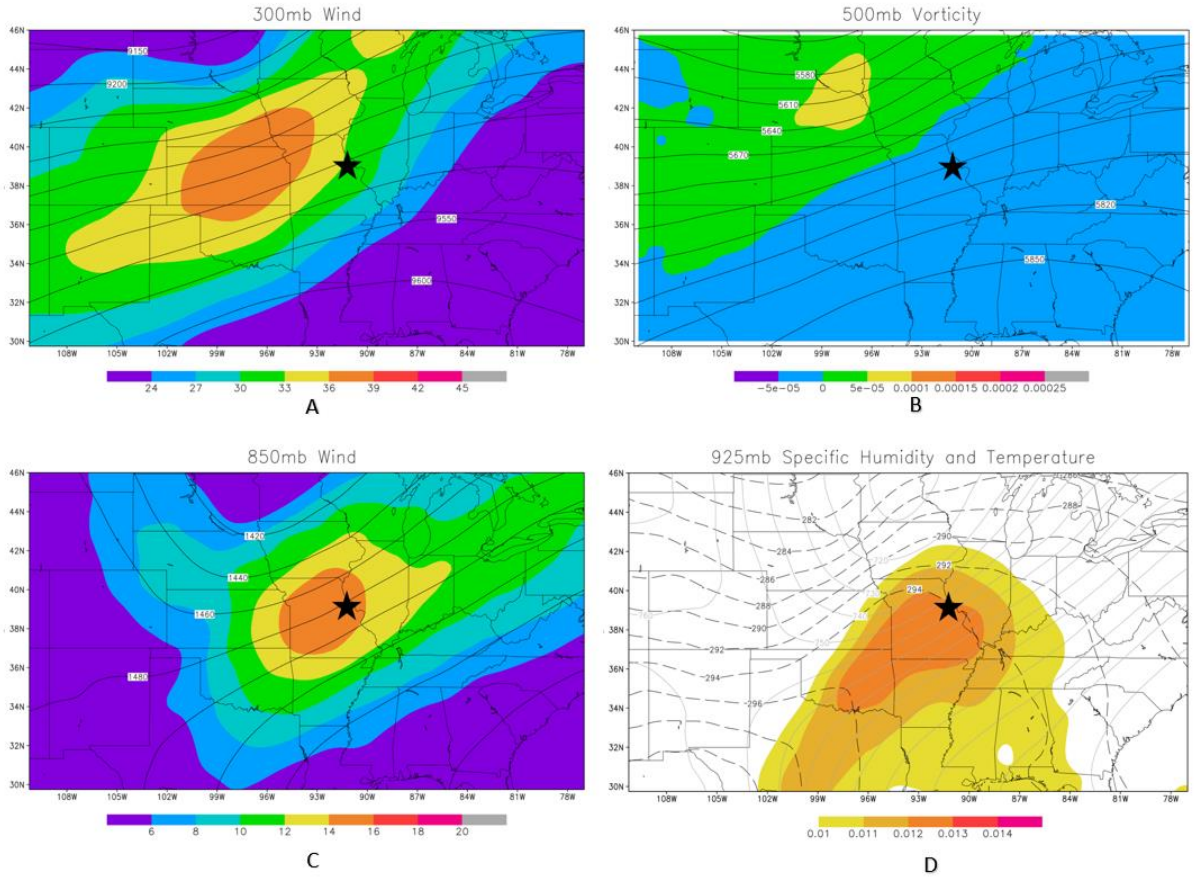

Figure 17 Composite maps from cluster 4. 300mb wind speed (m/s) (shaded) and geopotential height (m) (A), 500mb vorticity (s^{-2}) (shaded) and geopotential height (m) (B), 850mb wind speed (m/s) (shaded) and geopotential height (m) (C), 925mb specific humidity (kg/kg) (shaded), temperature (K) (black dashed), and geopotential height (m) (D). The black star represents the center of the TO.

One of the mesoscale features that is weaker than seen in other clusters are the CAPE values within the TO region. Values across the area range from 1500 J/kg up to 3000 J/kg (not shown). These CAPE values could be lower due to these TOs occurring across the Southeast and Midwest where they are not as high as those across the Great Plains (Shafer et al. 2009), (Anderson-Frey et al. 2018). CIN is present across the TO area but it is not abundant as values are between -10 J/kg and -20 J/kg. LCL values are closer to 900mb (not shown), which equates

to roughly 950 meters AGL. Figure 18b shows 0-1km SRH values between 120 m^2/s^2 and 160 m^2/s^2 across the TO region, which is enough to sustain an updraft within a mesocyclone. EBS values in Figure 18a show values above the 25-knot threshold defined by Thompson et al. (2007), with values between 35 knots and 47 knots. The SCP values for this cluster are the lowest among the clusters with values between 5 and 8 (not shown) across the TO region. While this is still above the threshold of 1, it is not as high as those in the other clusters. The same can be said for STP values. The STP values in Figure 18c only reaches 1 in a small area that is just SW of the TO region. According to Thompson et al. (2003(b)), this is not enough to be able to distinguish between tornadic and non-tornadic supercells. Figure 18d shows 0-1km EHI values above 1.5 and peaking somewhere between 2 and 2.5. With this, it indicates a larger probability for supercells to develop according to Rasmussen and Blanchard (1998). While tornadic supercells do occur within this cluster, they are not as easily as depicted given the various mesoscale features and composite indices. The ranking index for this cluster had a median value of 1.919, which is the highest value among all 4 clusters. The median value helps show that overall, this cluster contains some of the stronger outbreaks in the dataset. It also shows that on average, the TOs are becoming stronger as the median value for each cluster has grown as you progress from cluster 1 through cluster 4.

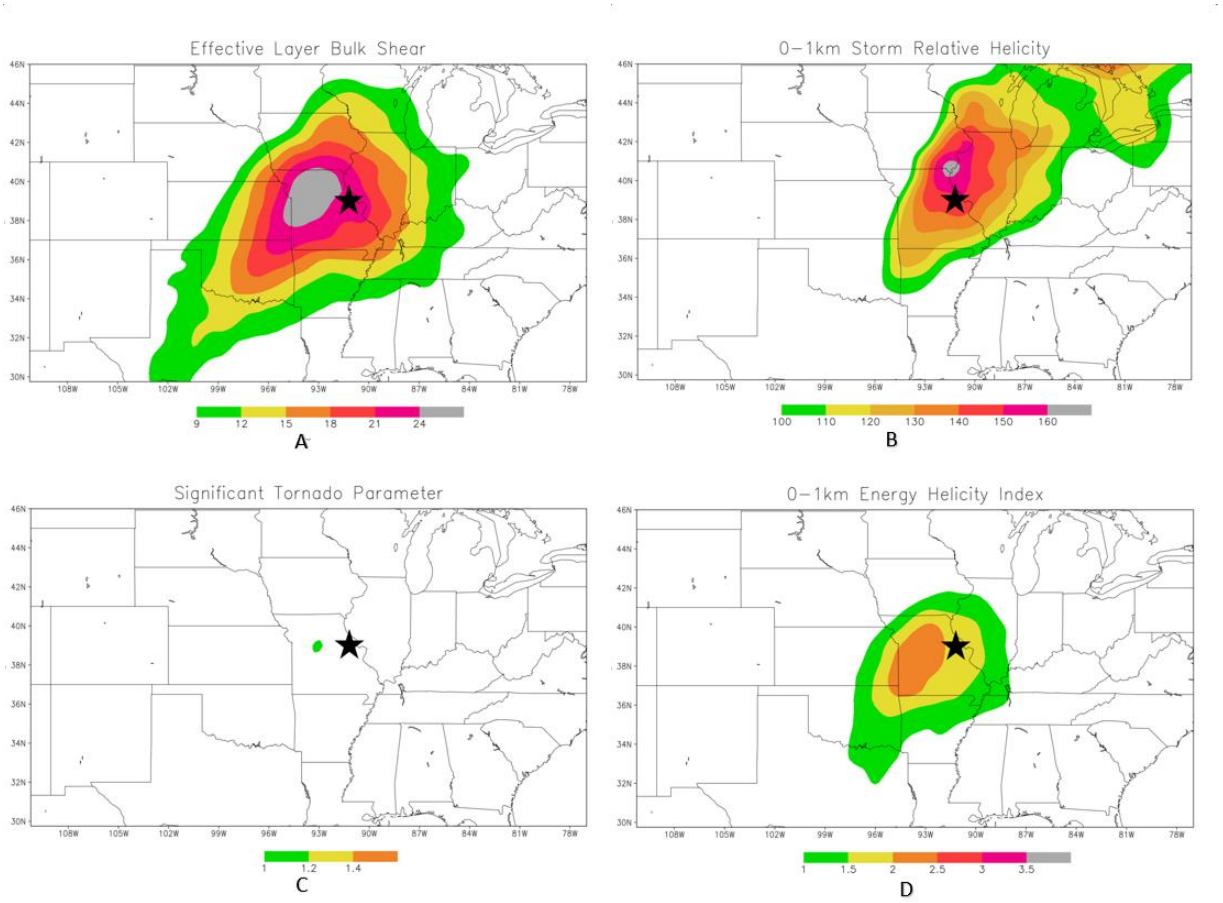

Figure 18 Composite maps from cluster 4. Effective layer bulk shear (m/s) (A), 0-1km storm relative helicity (m^2/s^2) (B), significant tornado parameter (C), 0-1km energy helicity index (D). The black star represents the center of the TO.

The example chosen for cluster 4 comes from April 15, 2011. The TO region is in the right exit region of the jet streak on the 300mb map in Figure 19a. This jet streak is stronger than the composite with values climbing above 45 m/s in the example whereas the composite only reaches up to 39 m/s. The 500mb in Figure 19b shows little to no CVA occurring, but the flow is much more meridional compared to the composite. The composite map in Figure 17a contains zonal flow due to the averaging done within the cluster that led to the zonal flow. The 850mb map in Figure 19c shows the LLJ being much stronger than the composite with values above 20

m/s across the TO region. The composite only contains values up to 16 m/s. The 925mb specific humidity values in Figure 19d is close to the same in the example as to the composite. There are a few areas with larger values but overall, it is of the same strength. The orientation of the specific humidity values also matches up well with those found in the composite map.

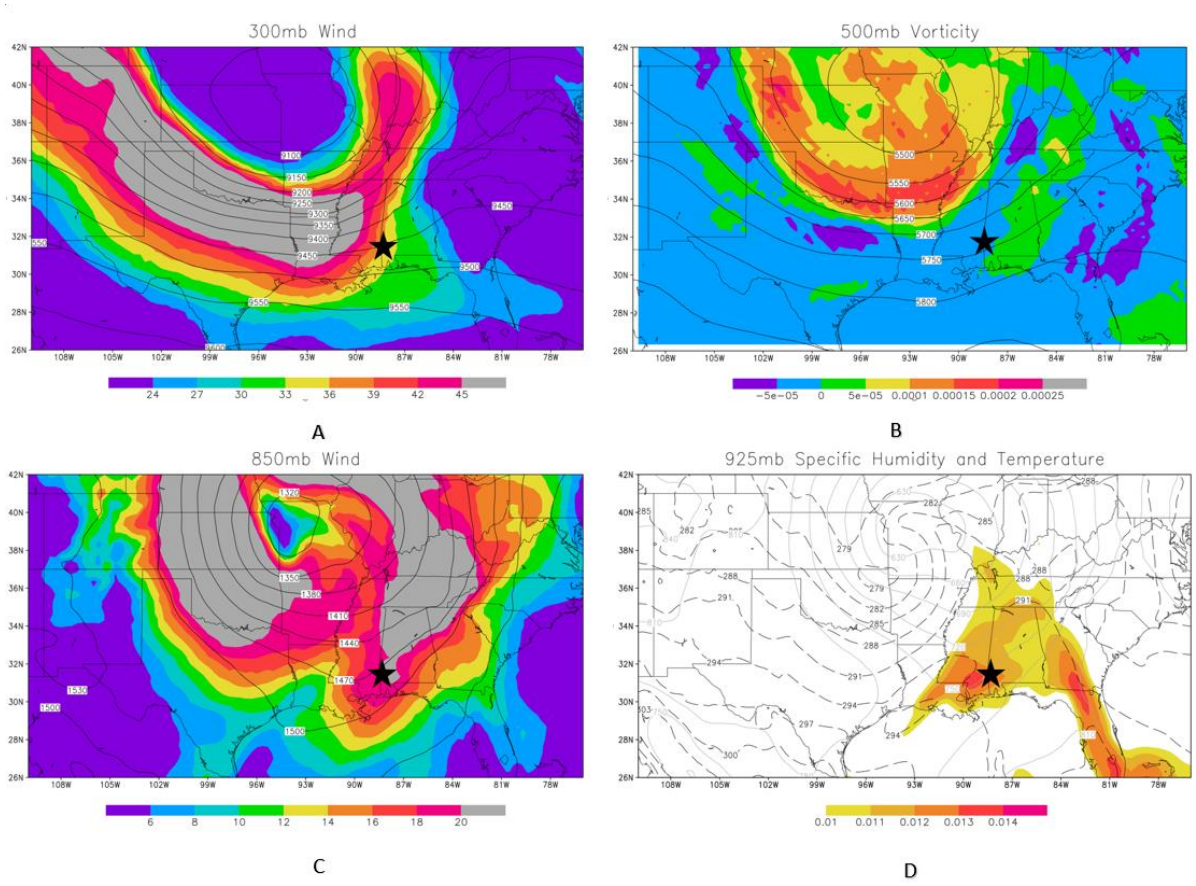

Figure 19 Maps from example outbreak from April 15, 2011. 300mb wind speed (m/s) (shaded) and geopotential height (m) (A), 500mb vorticity (s^{-2}) (shaded) and geopotential height (m) (B), 850mb wind speed (m/s) (shaded) and geopotential height (m) (C), and 925mb specific humidity (kg/kg) (shaded), temperature (K) (black dashed), geopotential height (m) (D). The black star represents the center of the TO.

CHAPTER V

DISCUSSION

The goal of this study was to characterize the environment of strong severe weather outbreaks that are tornado and hail dominant, thus more than likely supercell dominant. An unexpected result was also found to be that each cluster has a distinct time and location in which a TO occurred based upon the average month in which it occurred as well as the average latitude and longitude it occurred at. This could have an impact on the climatology of TOs and could possibly indicate a spatial and temporal change in TOs in the United States.

One of the first things that was noticed among the 300mb composite maps across all the clusters was that the TO region occurred within the right exit region of the jet streak. This finding is backed by Rose et al. (2004), but in their study they had found tornadoes occur 60% more often under the left exit region when compared to the right exit region. Another study from Clark et al. (2009) contained similar findings. They used the 4-quadrant model of a jet streak and placed storm reports in the quadrant they occurred in over an 11-year period. After April, the right exit region contained more storm reports than the left exit region. Clark et al. (2009) also explained that the increase in storm reports in the right exit region is explained by their composites of severe weather parameters. It showed that CAPE and 0-3km SRH were the highest in the right exit region and outweighed the effect of the upper-level convergence in this region. A study conducted by Gaffin and Parker (2006), found that most of the significant tornado events in the southern Appalachian region occur on the right side of the jet streak, including the

entrance and exit regions. While southern Appalachia is a small sample area for TOs, it still was able to provide evidence that the stronger tornado events occurred within the right side of the jet streak, including the exit region. In this study the composites were averages of the TOs that occurred within the cluster. For each cluster on average, the TO region occurred within the right exit region of the jet streak. It is also important to note that in this study, the definition of a TO was driven by the data itself and should produce a more accurate depiction of the TOs.

Other synoptic scale features such as CVA and WAA were found within each cluster, but of different magnitudes. The LLJ was also a prominent feature within each cluster and was typically co-located with a moisture plume at the 850mb level. Mesoscale features were also varied throughout each cluster, more specifically the values of STP and EHI. STP values were maximized in clusters 2 and 3 while the lowest STP value was recorded in cluster 4. STP values being the lowest in cluster 4 is an expected result as this is the most synoptically driven cluster among the four clusters. EHI values were the highest in cluster 2 which is expected because of this cluster contained the highest CAPE values as well as higher 0-1km SRH values. The lowest EHI values came from cluster 4 with this likely being due to CAPE being at its lowest among the composites within the four clusters. Since cluster 4 contained the strongest TOs among the four clusters and was the most synoptically driven, it is safe to assume that based on these results the strongest TOs typically occur when being forced by synoptic scale patterns.

One of the unforeseen findings of this paper is the possible change in the climatology of TOs across the United States. In this study, the average year in which a TO occurred changed for each cluster and kept progressing in time. The boxplot given in CHAPTER IVFigure 6 shows over time that cluster 4 is the favored set up for TOs. Another value that changed across the different clusters is the average location of each TO. Combining this with the map in CHAPTER

IVFigure 7 that shows the average TO location of each cluster, it can be inferred that the characteristics of TOs are shifting towards the setup within cluster 4 as well as moving eastward. The ranking index from Shafer and Doswell (2010) helps add on that while TOs are becoming more synoptically driven and moving further east, they are also becoming stronger as well.

The shift in TO frequency has possibly shifted to where more TOs are beginning to occur across the Southeastern United States and throughout the Ohio River Valley region and less are occurring across the Great Plains where the traditional Tornado Alley is located. This is backed by a study from Gensini and Brooks (2018) that shows tornado frequency increasing across the Southeast United States and the Ohio River Valley region and decreasing across the central and southern Great Plains regions. If this is the case, cluster 4 is a way to start to quantify just how these TOs are typically set up and how they tend to form. From the composite maps in cluster 4, there is not any evidence to suggest these TOs are result of a dry line event that is typically seen in the TOs across the southern Great Plains and seen in the composite maps in clusters 1 and 2. The orientation of the specific humidity values in CHAPTER IVFigure 17d would suggest it is more likely associated with a cold front rather than with a dryline. With an average time of outbreak in cluster 4 being in late April, most of the TOs are a result of an extratropical cyclone and the TO occurring within the warm sector before the cold front moves through. The presence of the LLJ is also important for transporting moisture and creating wind shear within the lower levels of the atmosphere.

CAPE values are lower in cluster 4 due to the location of the TOs. Most of the TOs in this cluster occur across the Southeast or Ohio River Valley and typically this is not an area that contains high CAPE values (Shafer et al. 2009), (Anderson-Frey et al. 2018). Springtime heating is enough to create temporary instability for the thunderstorms, but it is not able to build up the

way it does across the Great Plains. This is caused by upper-level temperatures being too warm to support large CAPE values in the Southeast whereas in the Great Plains the upper-level temperatures are much cooler. For cluster 4, and severe weather in the Southeast in general, wind shear now becomes a larger factor in the set-up of a TO. EBS and 0-1km SRH are important when looking for possible TO locations because they can show that wind shear is occurring at the lower levels and in the mid-levels of the atmosphere, while also being able to distinguish between supercells and non-supercells assuming certain thresholds are met.

There is not an exact formula for how a TO is set up but there are important combinations of synoptic and mesoscale ingredients that are needed to create enough instability and spin in the atmosphere to form tornadoes. It is important to understand TOs and how to quantify them to better forecast for these events and to make those that live in the path of these events aware of the severe weather that occurs around them.

CHAPTER VI

FUTURE RESEARCH

While this study tries to quantify TOs, it also revealed other trends that deserve to be investigated further. This includes the look into how climate change can be shifting the type of TO that is occurring across the United States and how those TOs in cluster 4 could be what we should expect to see more of in the future. Other factors that could help support such research could include the desertification of the Chihuahuan Desert in Northern Mexico and the deserts across western Texas. This is of importance because the dryline that forms across western portions of Oklahoma and Texas forms along the boundaries of the drier land. With the expansion of the drier land and desert areas, the origin of the dryline could be shifted to the east, pushing along with it bulk of the severe weather activity as the environment will be more supportive of the extreme weather east of the dryline. A publication produced by the Environmental Protection Agency (EPA 2016), indicates this is a possibility. Water resources in western Texas and Oklahoma come from what is known as the High Plains Aquifer System, which contains large amounts of groundwater that is used for irrigation of crops as well as usage by the public. This aquifer's groundwater has been rapidly depleted in areas across western Texas and Oklahoma since 1950 but has seen an increase in the rate of depletion since the 1970s. This leads to the disruption of water flow throughout the area and can lead to the drying up of rivers, streams, and lakes, according to Bartolino and Cunningham (2003). The drying of the lands water resources leads areas to become more arid and allows for vegetation to decrease. The

EPA concluded that the combination of the drier land, accompanied by warmer temperatures, could lead to the expansion of the Chihuahuan desert in northern Mexico and western Texas/Oklahoma. They also mentioned that wildfires and livestock grazing may accelerate the conversion of grassland to desert in response to the changes in the land surface. The expansion of this arid landscape could potentially change how the land heats up during the day and results in changes in differential heating across the landscape.

Further research to explain these changes could include the poleward expansion of the Tropics and further shifting tropical weather north. It would also be useful to update the TO database that was used for this study to include the most recent records of TO and to use the same data and methodologies within this study for the new database. This would allow for an additional 9 years of tornado records to be included within the study, further increasing the accuracy of the data.

REFERENCES

Anderson-Frey, A. K., Y. P. Richardson, A. R. Dean, R. L. Thompson, and B. T. Smith, 2018: Near-Storm Environments of Outbreak and Isolated Tornadoes. *Wea. Forecasting*, 33, 1397–1412, https://doi.org/10.1175/WAF-D-18-0057.1.

Bartolino, J.R. and W.L. Cunningham, 2003: Ground-Water Depletion Across the Nation. *United States Geological Survey (USGS)*. https://pubs.usgs.gov/fs/fs-103-03/JBartolinoFS(2.13.04).pdf.

Bell Laboratories, 2011: R version 2.13.0. Mississippi State University.

Clark, A.J., C.J. Schaffer, W.A. Gallus Jr., and K. Johnson-O'Mara, 2009: Climatology of Storm Reports Relative to Upper-Level Jet Streaks. *Wea. Forecasting*, 24, 1032 – 1051, https://doi.org/10.1175/2009WAF2222216.1.

Daoust, M, 2013: An analysis of tornado days in Missouri for the period 1950-2002. *Physical Geography*, 24, 467-487, https://doi.org/10.2747/0272-3646.24.6.467.

Davies, J. M., 2004: Estimations of CIN and LFC Associated with Tornadic and Nontornadic Supercells. *Wea. Forecasting*, 19, 714–726, https://doi.org/10.1175/1520-0434(2004)019<0714:EOCALA>2.0.CO;2.

Doswell, C. A., R. Edwards, R. L. Thompson, J. A. Hart, and K. C. Crosbie, 2006: A Simple and Flexible Method for Ranking Severe Weather Events. *Wea. Forecasting*, 21, 939–951, https://doi.org/10.1175/WAF959.1.

Doswell, C. A., and D. M. Schultz, 2006: On the use of indices and parameters in forecasting severe storms. *Electronic J. Severe Storms Meteor.*, 1, 1–22.

Environmental Protection Agency, 2016: What Climate Change Means for Texas. https://www.epa.gov/sites/production/files/2016-09/documents/climate-change-tx.pdf.

Fawbush, E. J., R. C. Miller, and L. G. Starrett, 1951: An Empirical Method of Forecasting Tornado Development. *Bull. Amer. Meteor. Soc.*, 32, 1–9, https://doi.org/10.1175/1520-0477-32.1.1.

Fuhrmann, C. M., C. E. Konrad, M. M. Kovach, J. T. McLeod, W. G. Schmitz, and P. G. Dixon, 2014: Ranking of Tornado Outbreaks across the United States and Their Climatological Characteristics. *Wea. Forecasting*, 29, 684–701, https://doi.org/10.1175/WAF-D-13-00128.1.

Gaffin, D.M. and S.S. Parker, 2006: A Climatology of Synoptic Conditions Associated with Significant Tornadoes across the Southern Appalachian Region. *Wea. Forecasting*, 21, 735-751, https://doi.org/10.1175/WAF951.1.

Galway, J.G., 1977: Some climatological aspects of tornado outbreaks. *Mon. Wea. Rev.*, 105, 477–484.

Gensini, V.A., Brooks, H.E. Spatial trends in United States tornado frequency. npj Clim Atmos Sci 1, 38 (2018). https://doi.org/10.1038/s41612-018-0048-2

Glickman, T., Ed., 2000: Glossary of Meteorology. 2nd ed. *Amer. Meteor. Soc*., 855 pp.

Grazulis, T. P., 1993: Significant Tornadoes 1680–1991: A Chronology and Analysis of Events. Environmental Films, 1340 pp.

Hart, J. A., and W. Korotky, 1991: The SHARP workstation vl.50 users guide. NOAA/National Weather Service. 30 pp.

Kelnosky, R.T., Tripoli, G.J. and Martin, J.E., 2018: Subtropical/polar jet influence on Plains and Southeast tornado outbreaks. *Nat Hazards*, 93, 373–392, https://doi.org/10.1007/s11069-018-3306-z.

Lee, C.C, 2011: Utilizing synoptic climatological methods to assess the impacts of climate change on future tornado-favorable environments. *Nat. Hazards*, 62, 325-343, https://doi.org/10.1007/s11069-011-9998-y.

Mercer, A. E., C. M. Shafer, C. A. Doswell, L. M. Leslie, and M. B. Richman, 2009: Objective Classification of Tornadic and Nontornadic Severe Weather Outbreaks. *Mon. Wea. Rev*., 137, 4355–4368, https://doi.org/10.1175/2009MWR2897.1.

Mercer, A., M. Richman, and L. Leslie, 2011: Identification of severe weather outbreaks using kernel principal component analysis., *Proc. Comp. Sci*., 6, 231-236. https://doi.org/10.1016/j.procs.2011.08.043

Mercer, A. E., C. M. Shafer, C. A. Doswell, L. M. Leslie, and M. B. Richman, 2012: Synoptic Composites of Tornadic and Nontornadic Outbreaks. *Mon. Wea. Rev*., 140, 2590–2608, https://doi.org/10.1175/MWR-D-12-00029.1.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc*., 87, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

Rasmussen, E. N., and D. O. Blanchard, 1998: A Baseline Climatology of Sounding-Derived Supercell andTornado Forecast Parameters. *Wea. Forecasting*, 13, 1148–1164, https://doi.org/10.1175/1520-0434(1998)013<1148:ABCOSD>2.0.CO;2.

Richman, M. B., 1986: Rotation of principal components. *J. Climatol.*, 6, 293–335.

Rasmussen, E. N., 2003: Refined Supercell and Tornado Forecast Parameters. *Wea. Forecasting*, 18, 530–535, https://doi.org/10.1175/1520-0434(2003)18<530:RSATFP>2.0.CO;2.

Rose, S. F., P. V. Hobbs, J. D. Locatelli, and M. T. Stoelinga, 2004: A 10-Yr Climatology Relating the Locations of Reported Tornadoes to the Quadrants of Upper-Level Jet Streaks. *Wea. Forecasting*, 19, 301–309, https://doi.org/10.1175/1520-0434(2004)019<0301:AYCRTL>2.0.CO;2.

Rousseeuw, P. J., 1987: Silhouettes: A graphical aid to the inter- pretation and validation for cluster analysis. *J. Comput. Appl. Math*., 20, 53–65.

Shafer, C. M., A. E. Mercer, C. A. Doswell, M. B. Richman, and L. M. Leslie, 2009: Evaluation of WRF Forecasts of Tornadic and Nontornadic Outbreaks When Initialized with Synoptic scale Input. *Mon. Wea. Rev*., 137, 1250–1271, https://doi.org/10.1175/2008MWR2597.1.

Shafer, C.M., and C.A. Doswell III, 2010: A multivariate index for ranking and classifying severe weather outbreaks. *Electronic J. Severe Storms Meteor*., 5, 1-39. https://www.researchgate.net/publication/228971917_A_multivariate_index_for_ranking_and_classifying_severe_weather_outbreaks.

Thompson, R. L., R. Edwards, and J. A. Hart, 2002: Evaluation and Interpretation of the Supercell Composite and Significant Tornado Parameters at the Storm Prediction Center. NOAA/NSSL/SPC. https://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_46942.htm

Thompson, R. L., C. M. Mead, and R. Edwards, 2003(a): Effective storm-relative helicity and bulk shear in supercell thunderstorm environments. Wea. Forecasting, 22, 102-115.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003(b): Close Proximity Soundings within Supercell Environments Obtained from the Rapid Update Cycle. *Wea. Forecasting*, 18, 1243–1261, https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

Thompson, R. L., C. M. Mead, and R. Edwards, 2004: Effective Bulk Shear in Supercell Thunderstorm Environments. Storm Prediction Center. https://www.spc.noaa.gov/publications/thompson/effshear.pdf

Thompson, R. L., C. M. Mead, and R. Edwards, 2007: Effective Storm-Relative Helicity and Bulk Shear in Supercell Thunderstorm Environments. *Wea. Forecasting*, 22, 102–115, https://doi.org/10.1175/WAF969.1.

Wallace, J. M. and P. V. Hobbs, 2006: *Atmospheric Science, an Introductory Survey*. Academic Press, 92, 345–346.

Weaver, S.J, S. Baxter, and A. Kumar, (2012): Climatic role of North American low-level jets on U.S. regional tornado activity. Journal of Climate, 25, 6666-6683. https://doi.org/10.1175/JCLI-D-11-00568.1.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 614 pp
Hidden test to allow template to find last page in document